

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 : Search time 12.1579 Seconds  
(without alignments)  
22.099 Million cell updates/sec

Title: US-09-833-079-2

Perfect score: 58  
Sequence: 1 AKFGMGAKKG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

al number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/6C.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfilest1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	65.5	106	3	US-08-963-168C-14
2	38	65.5	112	3	US-08-963-168C-16
3	38	65.5	116	3	US-08-963-168C-13
4	38	65.5	126	3	US-08-963-168C-7
5	38	65.5	132	3	US-08-963-168C-9
6	38	65.5	136	3	US-08-963-168C-6
7	36	62.1	111	3	US-08-963-168C-15
8	36	62.1	131	3	US-08-963-168C-8
9	35	60.3	129	4	US-09-108-020-41
10	35	60.3	660	3	US-09-111-085-2
11	35	60.3	660	4	US-09-376-781-5
12	35	60.3	946	3	US-08-560-005-4
13	35	60.3	946	4	US-09-418-540-4
14	33	56.9	141	4	US-09-091-725-51
15	33	56.9	334	4	US-09-232-191-5
16	33.5	56.0	334	4	US-09-232-200-5
17	33.5	56.0	334	4	US-09-232-197-5
18	32.5	56.0	334	4	US-09-232-201-5
19	32.5	56.0	597	4	US-09-232-191-23
20	32.5	56.0	597	4	US-09-232-191-37
21	32.5	56.0	597	4	US-09-232-200-23
22	32.5	56.0	597	4	US-09-232-200-91
23	32.5	56.0	597	4	US-09-232-200-99
24	32.5	56.0	597	4	US-09-232-197-23
25	32.5	56.0	597	4	US-09-232-197-91
26	32.5	56.0	597	4	US-09-232-197-99
27	32.5	56.0	597	4	US-09-232-201-23

28	32.5	56.0	597	4	US-09-232-201-91	Sequence 91, Appl
29	32.5	56.0	597	4	US-09-232-201-99	Sequence 99, Appl
30	32	55.2	70	4	US-08-847-065-16	Sequence 16, Appl
31	32	55.2	189	1	US-07-982-650D-2	Sequence 2, Appl
32	32	55.2	189	1	US-08-331-379-2	Sequence 2, Appl
33	32	55.2	196	1	US-07-982-650D-1	Sequence 1, Appl
34	32	55.2	196	1	US-07-982-650D-3	Sequence 3, Appl
35	32	55.2	196	1	US-07-982-650D-4	Sequence 4, Appl
36	32	55.2	196	1	US-07-982-650D-5	Sequence 5, Appl
37	32	55.2	196	1	US-08-331-379-1	Sequence 1, Appl
38	32	55.2	196	1	US-08-331-379-3	Sequence 3, Appl
39	32	55.2	196	1	US-08-331-379-4	Sequence 4, Appl
40	32	55.2	196	1	US-08-331-379-5	Sequence 5, Appl
41	32	55.2	226	3	US-09-195-286-1	Sequence 1, Appl
42	32	55.2	226	2	US-08-923-856-1	Sequence 1, Appl
43	32	55.2	299	3	US-09-216-234-1	Sequence 1, Appl
44	32	55.2	347	4	US-08-857-076-100	Sequence 100, App
45	32	55.2	420	4	US-08-847-065-19	Sequence 19, Appl

#### ALIGNMENTS

RESULT 1  
US-08-963-168C-14  
; Sequence 14, Application US/08963168C  
; Patent No. 6127166  
; GENERAL INFORMATION:  
; APPLICANT: Bayley, Hagan  
; APPLICANT: Cao, Qunping  
; APPLICANT: Wang, Yunjun  
; TITLE OF INVENTION: MOLLUSCAN LIGAMENT POLYPEPTIDES  
; TITLE OF INVENTION: AND GENES ENCODING THEM  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,168C  
; FILING DATE: 03-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/059001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 106 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-963-168C-14

Query Match 65.5%; Score 38; DB 3; Length 106;  
Best Local Similarity 77.8%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FGGMGAKKG 11  
DB 62 FGGMGAKKG 70

```
RESULT 2
US-08-963-168C-16
; Sequence 16, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quidping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 112 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-963-168C-16

Query Match          65.5%; Score 38; DB 3; Length 112;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

3 FCGMGAKKG 11
|||||
62 FCGMGGKGK 70

RESULT 3
US-08-963-168C-13
; Sequence 13, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quidping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
```

```
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 116 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-963-168C-13
```

```
Query Match          65.5%; Score 38; DB 3; Length 116;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
OY      3 FCGMGAKKG 11
|||||
Db       62 FCGMGGKGK 70
```

```
RESULT 4
US-08-963-168C-7
; Sequence 7, Application US/08963168C
; Patent No. 6127166
; GENERAL INFORMATION:
; APPLICANT: Bayley, Hagan
; APPLICANT: Cao, Quidping
; APPLICANT: Wang, Yunjaun
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES
; TITLE OF INVENTION: AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: IBM Compatible
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/963,168C
; FILING DATE: 03-NOV-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, Peter J.
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07917/059001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-08-963-168C-7
```

Query Match 65.5%; Score 38; DB 3; Length 126;  
Best Local Similarity 77.8%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKG 11  
|||||  
DB 82 FCGMGGKG 90

RESULT 5  
US-08-963-168C-9  
; Sequence 9, Application US/08963168C  
; Patent No. 6127166  
; GENERAL INFORMATION:  
; APPLICANT: Bayley, Hagan  
; APPLICANT: Cao, Quping  
; APPLICANT: Wang, Yunjaun  
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES  
; TITLE OF INVENTION: AND GENES ENCODING THEM  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,168C  
; FILING DATE: 03-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/059001  
; TELEPHONE: 617/542-5070  
; TELEFAX: 617/542-8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 132 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; FRAGMENT TYPE: Internal  
US-08-963-168C-9

Query Match 65.5%; Score 38; DB 3; Length 132;  
Best Local Similarity 77.8%; Pred. No. 18;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKG 11  
|||||  
DB 82 FCGMGGKG 90

RESULT 6  
US-08-963-168C-6  
; Sequence 6, Application US/08963168C  
; Patent No. 6127166  
; GENERAL INFORMATION:  
; APPLICANT: Bayley, Hagan  
; APPLICANT: Cao, Quping  
; APPLICANT: Wang, Yunjaun  
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES  
; TITLE OF INVENTION: AND GENES ENCODING THEM  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,168C  
; FILING DATE: 03-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/059001  
; TELECOMMUNICATION INFORMATION:

ADDRESSEE: Fish & Richardson P.C.  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/963,168C  
FILING DATE: 03-NOV-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Fasse, Peter J.  
REGISTRATION NUMBER: 32,983  
REFERENCE/DOCKET NUMBER: 07917/059001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 136 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: Internal  
FEATURE:  
OTHER INFORMATION: This translation is for SEQ ID NOS:1 & 2.

US-08-963-168C-6

Query Match 65.5%; Score 38; DB 3; Length 136;  
Best Local Similarity 77.8%; Pred. No. 19;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKG 11  
|||||  
DB 82 FCGMGGKG 90

RESULT 7  
US-08-963-168C-15  
; Sequence 15, Application US/08963168C  
; Patent No. 6127166  
; GENERAL INFORMATION:  
; APPLICANT: Bayley, Hagan  
; APPLICANT: Cao, Quping  
; APPLICANT: Wang, Yunjaun  
; TITLE OF INVENTION: MOLUSCAN LIGAMENT POLYPEPTIDES  
; TITLE OF INVENTION: AND GENES ENCODING THEM  
; NUMBER OF SEQUENCES: 43  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: Windows95  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/963,168C  
; FILING DATE: 03-NOV-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Fasse, Peter J.  
; REGISTRATION NUMBER: 32,983  
; REFERENCE/DOCKET NUMBER: 07917/059001  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070  
TELEFAX: 617/542-8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 111 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-963-168C-15

Query Match 62.1%; Score 36; DB 3; Length 111;  
Best Local Similarity 77.8%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMAKKG 11  
||| | |  
DB 58 FCGMAKKG 66

RESULT 8  
US-08-963-168C-8

Sequence 8, Application US/08963168C  
Patent No. 6127166

GENERAL INFORMATION:

APPLICANT: Bayley, Hagan

APPLICANT: Cao, Quiding

APPLICANT: Wang, Yunjuan

TITLE OF INVENTION: MOLLIUSCAN LIGAMENT POLYPEPTIDES

TITLE OF INVENTION: AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 43

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/963.168C

FILING DATE: 03-NOV-1997

ATTORNEY/AGENT INFORMATION:

NAME: Fasse, Peter J.

REGISTRATION NUMBER: 32,983

REFERENCE/DOCKET NUMBER: 07917/059001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617/542-5070

TELEFAX: 617/542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 131 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-08-963-168C-8

Query Match 62.1%; Score 36; DB 3; Length 131;  
Best Local Similarity 77.8%; Pred. No. 38;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMAKKG 11  
||| | |  
DB 78 FCGMAKKG 86

RESULT 9

US-09-108-020-41  
Sequence 41, Application US/09108020A  
Patent No. 6143561

GENERAL INFORMATION:

APPLICANT: Randall, Douglas D.

APPLICANT: Johnston, Mark L.

APPLICANT: Miernyk, Jan A.

APPLICANT: Luethy, Michael H.

APPLICANT: Mooney, Brian P.

TITLE OF INVENTION: USE OF DNA ENCODING PLASTID PYRUVATE DEHYDROGENASE AND

TITLE OF INVENTION: BRANCHED CHAIN OXOACID DEHYDROGENASE COMPONENTS TO

TITLE OF INVENTION: ENHANCE POLYHYDROXYALKANOATE BIOSYNTHESIS IN PLANTS

FILE REFERENCE: UWO 1482

CURRENT APPLICATION NUMBER: US/09/108.020A

EARLIER FILING DATE: 1998-06-30

EARLIER APPLICATION NUMBER: 60/051,291

EARLIER FILING DATE: 1997-06-30

EARLIER APPLICATION NUMBER: 60/055,255

EARLIER FILING DATE: 1997-08-01

EARLIER APPLICATION NUMBER: 60/076,544

EARLIER FILING DATE: 1998-03-02

NUMBER OF SEQ ID NOS: 54

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 41

LENGTH: 129

TYPE: PRT

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: Consensus

US-09-108-020-41

Query Match 60.3%; Score 35; DB 4; Length 129;  
Best Local Similarity 66.7%; Pred. No. 55;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 FCGMAKKG 11  
||| | |  
DB 36 FCGMAKKG 44

RESULT 10  
US-09-111-085-2

Sequence 2, Application US/09111085  
Patent No. 6100034

GENERAL INFORMATION:

APPLICANT: Stoye, Jonathan P

TITLE OF INVENTION: Detection of retroviral subtypes based upon envelope

FILE REFERENCE: 4238/75168

CURRENT APPLICATION NUMBER: US/09/111.085

EARLIER FILING DATE: 1998-07-07

EARLIER APPLICATION NUMBER: GB 9710154.7

EARLIER FILING DATE: 1997-05-16

NUMBER OF SEQ ID NOS: 16

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 660

TYPE: PRT

ORGANISM: Porcine retrovirus

US-09-111-085-2

Query Match 60.3%; Score 35; DB 3; Length 660;  
Best Local Similarity 66.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMAKKG 11  
:|| | |||  
DB 231 YGSGRKG 239

RESULT 11  
US-09-376-781-5

Sequence 5, Application US/09376781  
Patent No. 6261806  
GENERAL INFORMATION:  
APPLICANT: Banerjee, Papia T.  
APPLICANT: Patience, Clive  
APPLICANT: Andersson, Goran K.  
TITLE OF INVENTION: Molecular Sequence of Swine Retrovirus and Methods of  
Patent No. 6261806  
TITLE OF INVENTION: Use  
FILE REFERENCE: 61750-267  
CURRENT APPLICATION NUMBER: US/09/376,781  
CURRENT FILING DATE: 1999-08-18  
EARLIER APPLICATION NUMBER: 60/097,015  
EARLIER FILING DATE: 1998-08-18  
NUMBER OF SEQ ID NOS: 33  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 5  
LENGTH: 660  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: PERV-A  
OTHER INFORMATION: polypeptide sequence taken from Genbank Accession  
US-09-376-781-5  
Query Match 60.3%; Score 35; DB 4; Length 660;  
Best Local Similarity 66.7%; Pred. No. 2.5e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 3 FGGMGAKKG 11  
DB 231 YGGGSRKKG 239  
RESULT 12  
US-08-560-005-4  
Sequence 4, Application US/08560005  
Patent No. 6001354  
GENERAL INFORMATION:  
APPLICANT: Pot, David A.  
APPLICANT: Williams, Lewis T.  
APPLICANT: Jefferson, Anne Bennett  
APPLICANT: Majerus, Philip W.  
TITLE OF INVENTION: No. 6001354e1 Grb2 Associating Protein and Nucleic  
TITLE OF INVENTION: Acids Encoding Therefor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/560,005  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 2307K-0624000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 946 amino acids

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..946  
OTHER INFORMATION: /note= "ysc5ptase"  
US-08-560-005-4  
Query Match 60.3%; Score 35; DB 3; Length 946;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 3 FGGMGAKKG 11  
DB 644 FGGMASNRKG 652  
RESULT 13  
US-09-418-540-4  
Sequence 4, Application US/09418540  
Patent No. 6296848  
GENERAL INFORMATION:  
APPLICANT: Pot, David A.  
APPLICANT: Williams, Lewis T.  
APPLICANT: Jefferson, Anne Bennett  
APPLICANT: Majerus, Philip W.  
TITLE OF INVENTION: No. 6296848e1 Grb2 Associating Protein and Nucleic  
TITLE OF INVENTION: Acids Encoding Therefor  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew  
STREET: One Market Plaza, Stewart Tower, Suite 2000  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94105  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/418,540  
FILING DATE: 14-OCT-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/560,005  
FILING DATE: 17-NOV-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Dow, Karen B.  
REGISTRATION NUMBER: 29,684  
REFERENCE/DOCKET NUMBER: 2307K-0624000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-326-2400  
TELEFAX: 415-326-2422  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 946 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FEATURE:  
NAME/KEY: Region  
LOCATION: 1..946  
OTHER INFORMATION: /note= "ysc5ptase"  
US-09-418-540-4  
Query Match 60.3%; Score 35; DB 4; Length 946;  
Best Local Similarity 66.7%; Pred. No. 3.6e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 3 FGGMGAKKG 11

Db 644 FGGWASNGK 652

RESULT 14  
US-09-091-725-51

Sequence 51, Application US/09091725

Patent No. 6329141

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: Improved methods for transforming *Phaffia*  
TITLE OF INVENTION: and recombinant DNA for use therein

NUMBER OF SEQUENCES: 51

CORRESPONDENCE ADDRESS:

ADDRESSEE: Morrison & Foerster LLP

STREET: 2000 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: DC

COUNTRY: United States of America

ZIP: 20006-1888

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/091,725

FILING DATE: 23-DEC-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: EP 95203620.0

FILING DATE: 22-DEC-1995

APPLICATION NUMBER: EP 96200943.7

FILING DATE: 11-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: E. Victor Donahue

REGISTRATION NUMBER: 35,492

INFORMATION FOR SEQ ID NO: 51:

SEQUENCE CHARACTERISTICS:

LENGTH: 141 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-091-725-51

Query Match 56.9%; Score 33; DB 4; Length 141;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

2 KEGMGAK 9

Db 126 KEGMGAK 133

RESULT 15  
US-09-232-191-5

Sequence 5, Application US/09232191

Patent No. 6284487

GENERAL INFORMATION:

APPLICANT: Stahl, Andreas

APPLICANT: Hirsch, David J.

APPLICANT: Lodish, Harvey F.

TITLE OF INVENTION: Fatty Acid Transport Proteins

FILE REFERENCE: WHI97-21p3ME

CURRENT APPLICATION NUMBER: US/09/232,191

CURRENT FILING DATE: 1999-01-14

EARLIER APPLICATION NUMBER: 60/071,374

EARLIER FILING DATE: 1998-01-15

EARLIER APPLICATION NUMBER: 60/093,491

EARLIER FILING DATE: 1998-07-20

EARLIER APPLICATION NUMBER: 60/110,941

EARLIER FILING DATE: 1998-12-04

NUMBER OF SEQ ID NOS: 37

SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 5  
LENGTH: 334  
TYPE: PRT  
ORGANISM: Mycobacterium tuberculosis  
US-09-232-191-5

Query Match 56.0%; Score 32.5; DB 4; Length 334;  
Best Local Similarity 66.7%; Pred. No. 3.4e+02;  
Matches 8; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

QY 1 KEGMGAK-KG 11

Db 26 AVEGGMGLRLKG 37

Search completed: October 28, 2002, 17:24:31  
Job time : 13.1579 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 ; Search time 16.2105 Seconds  
(without alignments)  
65.203 Million cell updates/sec

Title: US-09-833-079-2  
Perfect score: 58  
Sequence: 1 AKFGMGAKKG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	67.2	237	2	hypothetical prote
2	65.5	156	2	probable bzip tran
3	65.5	516	2	probable phosphati
4	65.5	944	2	hypothetical prote
5	63.8	17	2	17k basolateral pl
6	63.8	113	1	hemerythrin - sipu
7	63.8	142	2	polyribonucleotide
8	63.8	142	2	polyribonucleotide
9	63.8	371	2	acyl-CoA dehydroge
10	63.8	374	2	heat shock protein
11	63.8	374	2	molecular chaperon
12	63.8	575	2	RNA helicase - fru
13	62.1	185	1	fibrillar protein p
14	62.1	386	1	hydrogenase (EC 1.
15	62.1	387	2	probable amidohydr
16	62.1	399	2	probable adenosylh
17	62.1	1957	2	myosin heavy chain
18	62.1	1957	2	skeletal myosin -
19	60.3	88	1	ribosomal protein
20	60.3	140	2	ribosomal protein
21	60.3	182	2	asparaginase famli
22	60.3	292	2	asparaginase (agmat
23	60.3	305	2	probable hycd prot
24	60.3	316	2	DNA repair protein
25	60.3	350	2	conserved hypotnet
26	60.3	579	2	hypothetical prote
27	60.3	604	2	probable homoaconi
28	60.3	721	2	probable tpr prote
29	60.3	756	2	F71340

30	60.3	758	2	F71301	probable tpr prote
31	60.3	762	2	C71340	probable tpr prote
32	60.3	853	2	AB2020	hypothetical prote
33	60.3	909	2	S32538	cGMP-gated cation
34	60.3	946	2	S48433	inositol-1,4,5-tri
35	60.3	1829	2	T34239	hypothetical prote
36	60.3	3512	2	T17121	CpY protein - midg
37	58.6	87	2	JC5035	hypothetical prote
38	58.6	88	2	T43610	probable IS1617 tr
39	58.6	128	2	C53380	polyribonucleotide
40	58.6	137	2	E83659	hypothetical prote
41	58.6	142	2	T52145	ribosomal protein
42	58.6	142	2	S67619	ribosomal protein
43	58.6	144	1	HSUR2P	histone H2b.1, spe
44	58.6	145	1	R3YL16	ribosomal protein
45	58.6	145	2	S41193	ribosomal protein

## ALIGNMENTS

RESULT 1  
A83541  
hypothetical protein PA0833 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: A83541  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapid, K.; L  
Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:2043737  
A:Accession: A83541  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-237 <SNO>  
A:Cross-references: GB:AE004518; GB:AE004091; NID:g9946725; PIDN:AA04222.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA0833

Query Match  
Best Local Similarity 67.2% Score 39; DB 2; Length 237;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AKFGMGAKKG 11  
||:|||||  
DB 42 AKYGGGALAG 52

RESULT 2  
C84556  
probable bzip transcription factor [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: C84556  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y  
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon  
euss, D.; Nieman, W.C.; White, O.; Elsen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487  
A:Accession: C84556  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-156 <SNO>  
A:Cross-references: GB:AE002093; NID:g6598809; PIDN:AA18682.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2g17770  
A:Map position: 2

Query Match 65.5% Score 38; DB 2; Length 156;

Best Local Similarity 63.6%; Pred. No. 19;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 1 AKFGMGAKKG 11  
|||||:|:|:  
Db 106 AKFGCLGKKG 116

## RESULT 3

T50190  
Probable phosphatidylserine decarboxylase proenzyme 1 precursor [imported] - fission yeast  
C:Species: Schizosaccharomyces pombe  
C>Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jun-2000  
C:Accession: T50190  
R:Barrell, B.G.; Rajandream, M.A.; McDougall, R.C.; McLean, J.; Harris, D.  
Submitted to the EMBL Data Library, November 1999  
A:Reference number: 225045  
A:Accession: T50190  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-516 <BAR>  
A:Cross-references: EMBL:AL133225; PIDN:CA861769.1; GSPDB:GN00066; SPDB:SPAC25B8.03  
A:Experimental source: strain 972h(-); cosmid c25B8  
C:Genetics: SPDB:SPAC25B8.03  
A:Gene: SPDB:SPAC25B8.03  
A:Map position: 1

Query Match 65.5%; Score 38; DB 2; Length 516;  
Best Local Similarity 77.8%; Pred. No. 57;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 3 FGGMGAKKG 11  
|||:|:|:  
Db 22 FGGVGAKKG 30

## RESULT 4

D82926  
Hypothetical protein U0166 [imported] - Ureaplasma urealyticum  
C:Species: Ureaplasma urealyticum  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 20-Aug-2000  
C:Accession: D82926  
R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.  
Submitted to GenBank, February 2000  
A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a mit  
A:Reference number: A82870  
A:Accession: D82926  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-944 <GLA>  
A:Cross-references: GB:AE002116; GB:AF222894; NID:96899118; PIDN:AAF30573.1; GSPDB:GN001  
A:Experimental source: serovar 3; biovar 1  
C:Genetics:  
A:Gene: U0166  
A:Genetic code: SGC3

Query Match 65.5%; Score 38; DB 2; Length 944;  
Best Local Similarity 70.0%; Pred. No. 98;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 2 KFGMGAKKG 11  
|||:|:|:  
Db 416 KLGCTGAKG 425

## RESULT 5

A61019  
17K basolateral plasma membrane protein Proct17 - rat (fragment)  
N:Alternate names: Proct17  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 03-May-1994  
C:Accession: A61019  
R:Schlecht, H.

Histochemistry 93, 513-518, 1990  
A:Title: N-terminal amino acid sequence, immunohistochemical localization and tissue  
A:Reference number: A61019; MUID:90236771  
A:Accession: A61019  
A:Molecule type: protein  
A:Residues: 1-17 <SCH>  
C:Keywords: Intestine; membrane protein

Query Match 63.8%; Score 37; DB 2; Length 17;  
Best Local Similarity 60.0%; Pred. No. 3.8;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;  
OY 2 KFGMGAKKG 11  
::|||:|:|:  
Db 7 EYGGKGGKKG 16

## RESULT 6

HRIN  
hemerythrin - sipunculid (Siphonosoma cumense)  
C:Species: Siphonosoma cumense  
C>Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 01-Aug-1997  
C:Accession: JF0556  
R:Uchida, T.; Yano, H.; Satake, K.; Kubota, I.; Tsugita, A.  
Protein Seq. Data Anal. 3, 141-147, 1990  
A:Title: The amino acid sequence of hemerythrin from Siphonosoma cumense.  
A:Reference number: JF0556; MUID:90301732  
A:Accession: JF0556  
A:Molecule type: protein  
A:Residues: 1-113 <UCH>  
A:Note: 3-Glu, 10-Asp, 60-Gly, 66-Asn, and 83-Gln were also found  
C:Comment: Hemerythrin is a respiratory protein found in several phyla of marine invertebrates.  
C:Superfamily: hemerythrin  
C:Keywords: iron; oxygen carrier  
F:25,54,58,73,77,101,106/Binding site: 2Fe-O cluster (His, His, Glu, His, His, His, A

Query Match 63.8%; Score 37; DB 1; Length 113;  
Best Local Similarity 70.0%; Pred. No. 21;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 1 AKFGMGAKK 10  
|||:|:|:  
Db 65 AKYGGYGAHK 74

## RESULT 7

AC1102  
polyribonucleotide nucleotidyltransferase domain present [imported] - Listeria monocytogenes  
C:Species: Listeria monocytogenes  
C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AC1102  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baguero, F.; Berche, P.; Bloec  
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussureget, O.; Entian, K.D.; Fshih,  
D.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkát, G.; Madueno, E.; Maltournam, A.;  
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla  
A:Title: Comparative genomics of Listeria species.  
A:Reference number: AB1077; MUID:21537279; PMID:11679669  
A:Accession: AC1102  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-142 <GLA>  
A:Cross-references: GB:NC\_003210; PIDN:CAD00745.1; PID:q16409583; GSPDB:GN00177  
A:Experimental source: strain EGD-e  
C:Genetics:  
A:Gene: lmo0218  
C:Superfamily: polyribonucleotide nucleotidyltransferase homolog yabr

Query Match 63.8%; Score 37; DB 2; Length 142;  
Best Local Similarity 63.6%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;



QY 1 AKFGMGAKKG 11  
Db 132 SKRGGRGAKKG 142

## RESULT 8

AC1464  
polyribonucleotide nucleotidyltransferase domain present [imported] - *Listeria innocua*  
C:Species: *Listeria innocua*  
C:Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 14-Dec-2001  
C:Accession: AC1464  
R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fehli, H.; Jones, L.M.; Karst, U.  
Science 294, 849-852, 2001  
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkut, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schluter, T.; Simoes, N.; Tixeront, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.  
Reference number: AB1077; MUID:21537279; PMID:11679669  
C:Accession: AC1464  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-142 <GIA>  
A:Cross-references: GB:AL592022; PIDN:CAC95483.1; PID:916412679; GSPDB:GN00178  
A:Experimental source: strain Clp11262  
C:Genetics:  
A:Gene: Lin0250  
C:Superfamily: polyribonucleotide nucleotidyltransferase homolog yabr

Query Match 63.8%; Score 37; DB 2; Length 142;  
Best Local Similarity 63.8%; Pred. No. 26;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 AKFGMGAKKG 11  
Db 132 SKRGGRGAKKG 142

## RESULT 9

E90373  
acyl-CoA dehydrogenase (acd-3) [imported] - *Sulfolobus solfataricus*  
C:Species: *Sulfolobus solfataricus*  
C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 15-Jun-2001  
C:Accession: E90373  
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.U.; Chan, J.; Jett, R.A.; Ragan, M.A.; Jensen, C.W.; Van der Oost, J.  
Submitted to Genbank, April 2001  
Description: *Sulfolobus solfataricus* complete genome.  
A:Reference number: A99139  
A:Accession: E90373  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-371 <KUR>  
A:Cross-references: GB:AE006641; NID:913815346; PIDN:AAK42244.1; GSPDB:GN00155  
C:Genetics:  
A:Gene: acd-3  
C:Superfamily: acyl-CoA dehydrogenase

Query Match 63.8%; Score 37; DB 2; Length 371;  
Best Local Similarity 77.8%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 FGGMGAKKG 11  
Db 324 FGGMGAKKG 332

## RESULT 10

S41758  
heat shock protein dnaJ - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*

C:Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 20-Aug-1999  
C:Accession: S41758; A41873  
R:Behrens, S.; Narberhaus, F.; Bahl, H.  
FEMS Microbiol. Lett. 114, 53-60, 1993  
A:Title: Cloning, nucleotide sequence and structural analysis of the *Clostridium acetobutylicum* dnaJ gene  
A:Reference number: S41758; MUID:94123950  
A:Accession: S41758  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-374 <BEH>  
A:Cross-references: EMBL:X69050; NID:91246429; PIDN:CAA48792.1; PID:9433079  
R:Narberhaus, F.; Giebler, K.; Bahl, H.  
J. Bacteriol. 174, 3290-3299, 1992  
A:Title: Molecular characterization of the dnaJ gene region of *Clostridium acetobutylicum*  
A:Reference number: A41873; MUID:92250425  
A:Accession: A41873  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-72 <NAR>  
A:Cross-references: GB:M74569; NID:9144828; PIDN:AAA23247.1; PID:9144832  
C:Genetics:

A:Gene: dnaJ  
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology  
C:Keywords: DNA replication; heat shock; molecular chaperone  
F:5-70/Domain: dnaJ amino-terminal homology <DNJ>  
F:150-157/Region: CXXCXGKG repeat  
F:167-174/Region: CXXCXGKG repeat  
F:189-200/Region: CXXCXGKG repeat  
F:207-214/Region: CXXCXGKG repeat

Query Match 63.8%; Score 37; DB 2; Length 374;  
Best Local Similarity 87.5%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGMGAKKG 11  
Db 154 GGTGAKKG 161

## RESULT 11

C97058  
molecular chaperones DnaJ (HSP40 family) [imported] - *Clostridium acetobutylicum*  
C:Species: *Clostridium acetobutylicum*  
C:Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 30-Sep-2001  
C:Accession: C97058  
R:Moiling, J.; Britton, G.; Omeletchenko, M.V.; Matkova, K.S.; Zeng, Q.; Gibson, R.; L.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium  
A:Reference number: A96900; MUID:21359325; PMID:21359325  
A:Accession: C97058  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-374 <KUR>  
A:Cross-references: GB:AE001437; PIDN:AAK79254.1; PID:915024211; GSPDB:GN00168  
A:Experimental source: *Clostridium acetobutylicum* ATCC624  
C:Genetics:  
A:Gene: CAC1283  
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

Query Match 63.8%; Score 37; DB 2; Length 374;  
Best Local Similarity 87.5%; Pred. No. 63;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGMGAKKG 11  
Db 154 GGTGAKKG 161

## RESULT 12

S11485  
RNA helicase - fruit fly (*Drosophila melanogaster*)  
C:Species: *Drosophila melanogaster*

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-Jan-2001  
C:Accession: S11485  
R:Dorer, D.R.; Christensen, A.C.; Johnson, D.H.  
Nucleic Acids Res. 18, 5489-5494, 1990  
A>Title: A novel RNA helicase gene tightly linked to the Triplo-lethal locus of Drosophila  
A:Reference number: S11485; MUID:91016833  
A:Accession: S11485  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-575 <DOR>  
A:Cross-references: EMBL:X52846; NID:g8443; PIDN:CAA37037.1; PID:g8444  
C:Genetics:  
A:Gene: FlyBase:Rm62  
A:Cross-references: FlyBase:FBgn0003261  
C:Superfamily: ATP-dependent RNA helicase DBP1  
C:Keywords: ATP; nucleotide-binding; P-loop  
F:181-188/Region: nucleotide-binding motif A (P-loop)  
F:287-292/Region: nucleotide-binding motif B  
F:291-294/Region: DEAD motif

Query Match 63.8%; Score 37; DB 1; Length 575;  
Best Local Similarity 70.0%; Pred. No. 94;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 KFGMGAKKG 11  
| | | | |  
Db 542 RFGGGRKKG 551

RESULT 13  
YOECP  
fimbrial protein papa precursor - Escherichia coli  
N:Alternate names: pap p11  
C:Species: Escherichia coli  
C>Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C:Accession: A23221; S25216; A05229; S16395  
R:Bagu, M.; Normark, S.; Hardy, J.; O'Hanley, P.; Lark, D.; Olsson, O.; Schoolnik, G.; F  
J. Bacteriol. 157, 330-333, 1984  
A>Title: Nucleotide sequence of the papa gene encoding the pap pilus subunit of human ur  
A:Reference number: A91794; MUID:84087728  
A:Accession: A23221  
A:Molecule type: DNA  
A:Residues: 1-165 <BAG>  
A:Cross-references: GB:X03391; GB:K01116; GB:X03392; NID:g42309; PIDN:CAA27126.1; PID:g4  
M. Marlund, B.I.; Tennent, J.M.; Garcia, E.; Hamers, A.; Bagu, M.; Lindberg, F.; Gastra  
Mol. Microbiol. 6, 2225-2242, 1992  
A>Title: Horizontal gene transfer of the Escherichia coli pap and prs pil operons as a  
A:Reference number: S25205; MUID:93023852  
A:Accession: S25216  
A:Molecule type: DNA  
A:Residues: 1-165 <MA2>  
A:Cross-references: EMBL:X61239; NID:g42290; PIDN:CAA43562.1; PID:g42293  
A:Experimental source: strain J96  
C:Genetics:  
A:Gene: papa  
C:Superfamily: F7-2 fimbrial protein  
C:Keywords: fimbria  
F:1-23/Domain: signal sequence #status predicted <Sig>  
F:23-185/Product: fimbrial protein papa #status predicted <Mat>

Query Match 62.1%; Score 36; DB 1; Length 185;  
Best Local Similarity 87.5%; Pred. No. 50;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 GGMGAKKG 11  
| | | | |  
Db 90 GGMGAKKG 97

RESULT 14  
S33852  
hydrogenase (EC 1.18.99.1) (Nife) hyda - Wolinella succinogenes  
C:Species: Wolinella succinogenes

C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 19-May-2000  
C:Accession: S33852; S22404; S33063  
R:Dross, F.; Geisler, V.; Lenger, R.; Theis, F.; Kraft, T.; Fahrenholz, F.; Kojro, E  
Eur. J. Biochem. 214, 949-950, 1993  
A>Title: Correction: The quinone-reactive Ni/Fe-hydrogenase of Wolinella succinogenes  
A:Reference number: S33852; MUID:93307313  
A:Accession: S33852  
A:Molecule type: DNA  
A:Residues: 1-386 <DRO>  
A:Cross-references: EMBL:X65189; NID:g296081; PIDN:CAA46302.1; PID:g296082  
A>Note: this is a revision to the sequence from reference S22404  
R:Dross, F.; Geisler, V.; Lenger, R.; Theis, F.; Kraft, T.; Fahrenholz, F.; Kojro, E  
Eur. J. Biochem. 206, 93-102, 1992  
A>Title: The quinone-reactive Ni/Fe-hydrogenase of Wolinella succinogenes.  
A:Reference number: S22404; MUID:92267032  
A:Accession: S22404  
A:Molecule type: DNA  
A:Residues: 33-86, 'VVA', 90-91, 'PKNRMSGTRVS', 105-106, 'F', 109-215, 'LDVS', 221-227, 'KL', 2  
A:Cross-references: EMBL:X65189  
A>Note: this sequence has been revised in reference S33852  
C:Genetics:  
A:Gene: hyda  
C:Superfamily: hydrogenase (Nife) small chain  
C:Keywords: 4Fe-4S; metalloprotein; oxidoreductase  
F:85,88,185,218/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted  
F:256,259,284,290/Binding site: 4Fe-4S cluster (His, Cys, Cys, Cys) (covalent) (type

Query Match 62.1%; Score 36; DB 1; Length 386;  
Best Local Similarity 70.0%; Pred. No. 98;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 KFGMGAKKG 11  
| | | | |  
Db 273 EFGDEGAKKG 282

RESULT 15  
G71097  
probable amidohydrolase - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
C:Accession: G71097  
R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu  
DNA Res. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophil  
A:Reference number: A71000; MUID:98344137  
A:Accession: G71097  
A>Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-387 <KAW>  
A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30141.1; PID:g3257458  
A:Experimental source: strain OT3  
A>Note: this accession replaces an interim accession for a sequence replaced by GenBa  
C:Genetics:  
A:Gene: PH1043  
C:Superfamily: hipurate hydrolase

Query Match 62.1%; Score 36; DB 2; Length 387;  
Best Local Similarity 70.0%; Pred. No. 98;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 AKFGMGAKK 10  
| | | | |  
Db 136 AEEGGLGAKK 145

Search completed: October 28, 2002, 17:23:15  
Job time: 18.2105 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:15 : Search time 8.10526 Seconds

(without alignments)  
52.548 Million cell updates/sec

Title: US-09-833-079-2  
Perfect score: 58  
Sequence: 1 AKFGMGAKKG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapept 0.5

Searched: 105224 seqs, 38719550 residues  
Minimum number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SWISSPROT\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38	65.5	944	1 Y166_UREPA	Q9Pqx7 ureaplasma
2	37	63.8	113	1 HEWT_SIPCU	P22766 siphonoma
3	37	63.8	374	1 DNAM2_CIOAB	P30725 clostridium
4	37	63.8	575	1 RM62_DROME	P19109 drosophila
5	37	63.8	989	1 DUP4_HUMAN	O9Y2h0 homo sapien
6	36	62.1	185	1 PAPA_ECOLI	P04127 escherichia
7	36	62.1	354	1 MBHS_WOLSV	P31884 wolfinella s
8	36	62.1	992	1 DLP4_RAT	P97839 rattus norv
9	35	60.3	88	1 RS19_MYCCA	P10132 mycoplasma
10	35	60.3	140	1 RL3_PJARO	P72233 planobispor
11	35	60.3	182	1 AHT1_YEAST	P23589 saccharomyc
12	35	60.3	350	1 RA55_SCHPO	O14129 schizosacch
13	35	60.3	648	1 CH6B_DROME	O9Ypss drosophila
14	35	60.3	909	1 CNG4_HUMAN	O14028 homo sapien
15	35	60.3	946	1 Y1A2_YEAST	P40559 saccharomyc
16	35	60.3	952	1 Y014_BACAN	O9Yn18 bacillus an
17	34	58.6	128	1 YABR_BACSD	P37560 bacillus su
18	34	58.6	142	1 RS16_CANAL	O94017 candida alb
19	34	58.6	142	1 RS16_TORRU	O9Yex7 tortula rur
20	34	58.6	142	1 RS16_YEAST	P40213 saccharomyc
21	34	58.6	144	1 H2B1_PARAN	P02290 parechnus
22	34	58.6	145	1 RS16_FRING	O22647 fritillaria
23	34	58.6	145	1 RS16_GOSHI	P46293 gossypium h
24	34	58.6	145	1 RS16_LUPPO	P16149 lupinus pol
25	34	58.6	146	1 RS16_ARATH	Q42340 arabidopsi
26	34	58.6	164	1 DH1B_ORYSA	P22911 oryza sativ
27	34	58.6	164	1 DH1C_ORYSA	P22912 escherichia
28	34	58.6	259	1 SRUD_ECOLI	P05707 escherichia
29	34	58.6	273	1 MPT4_YEAST	P33015 saccharomyc
30	34	58.6	353	1 HN3G_MOUSE	P33584 mus musculu
31	34	58.6	354	1 HN3G_RAT	P33183 rattus norv
32	34	58.6	417	1 DNJH_ATRND	P43644 atrophic nu
33	34	58.6	419	1 HFLK_ECOLI	P25662 escherichia

34	34	58.6	428	1 PUR2_BACHD	Q9Kf52 bacillus ha
35	34	58.6	1221	1 TOP2_TRYBB	P12531 trypanosoma
36	34	58.6	1267	1 DHRL_YEAST	O04217 saccharomyc
37	34	58.6	1723	1 PM20_CHLPP	O92812 chlamydia p
38	33	56.9	140	1 RS16_SCHPO	O60144 schizosacch
39	33	56.9	144	1 RS16_CAEEL	Q22054 caenorhabd
40	33	56.9	144	1 RS16_MOUSE	P14131 mus musculu
41	33	56.9	145	1 RS16_HUMAN	P17008 homo sapien
42	33	56.9	226	1 TPIS_RHIER	P96985 rhizobium e
43	33	56.9	382	1 DNJ1_HAEIN	P43735 haemophilus
44	33	56.9	395	1 HFLK_VIBCH	O9Kv09 vibrio chol
45	33	56.9	475	1 Z131_HUMAN	P52739 homo sapien

## ALIGNMENTS

RESULT 1					
ID	Y166_UREPA	STANDARD:	PRT:	944 AA.	
AC	O9Pqx7:				
DT	16-OCT-2001 (Rel. 40, Created)				
DT	16-OCT-2001 (Rel. 40, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Hypothetical protein U0166.				
GN	U0166.				
OS	Ureaplasma parvum (Ureaplasma urealyticum biotype 1).				
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;				
OC	Mycoplasmataceae; Ureaplasma.				
OX	NCBI_TaxID=134821;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SEVOVAR 3;				
RX	MEDLINE=20500219; Pubmed=11048724;				
RA	Glass J.I., Lefkowitz E.J., Glass J.S., Helner C.R., Chen E.Y.,				
RA	Cassell G.H.;				
RT	"The complete sequence of the mucosal pathogen Ureaplasma				
RT	urealyticum".				
RL	Nature 407:757-762(2000).				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>				
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).				
CC	-----				
DR	EMBL: AE002116; AAF30573.1; -				
KW	Hypothetical protein; Complete proteome.				
SQ	SEQUENCE 944 AA; 107923 MW; B5755E0CC99B8274 CRC64;				
Query Match					
Best local Similarity 70.0%; Pred. No. 40;					
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;					
QY	2 KFGMGAKKG 11				
DB	416 KLGTCGAKG 425				
RESULT 2					
ID	HEWT_SIPCU	STANDARD:	PRT:	113 AA.	
AC	P22766:				
DT	01-AUG-1991 (Rel. 19, Created)				
DT	01-AUG-1991 (Rel. 19, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	Hemerythrin.				
OS	Siphonoma cumanense.				
OC	Eukaryota; Metazoa; Siphunculida; Siphunculidae; Siphonoma.				
OX	NCBI_TaxID=6444;				
RN	[1]				

RP SEQUENCE.  
RX MEDLINE=90301732; PubMed=2362933;  
RA Uchida T., Yano H., Satake K., Kibota I., Tsugita A.;  
RT "The amino acid sequence of hemerythrin from Siphonosoma cumanense";  
RL Protein Seq. Data Anal. 3:141-147(1990).  
CC -1- FUNCTION: HEMERYTHRIN IS A RESPIRATORY PROTEIN IN BLOOD CELLS OF  
CERTAIN MARINE WORMS. THE OXYGEN-BINDING SITE IN EACH CHAIN  
CONTAINS TWO IRON ATOMS.  
CC -1- SUBUNIT: HOMOTRIMER.  
CC -1- SIMILARITY: BELONGS TO THE HEMERYTHRIN FAMILY.  
DR HSP; J0556; HRIN.  
DR HSP; P02246; ZMO.  
DR InterPro: IPR002063; Hemerythrin.  
DR Pfam: PF01814; Hemerythrin; 1.  
DR PRINTS: PR00186; HEMERYTHRIN.  
DR PROSITE: PS00550; HEMERYTHRINS; 1.  
KW Oxygen transport; Metal-binding; Iron.  
FT METAL 25 25 IRON 1 (BY SIMILARITY).  
FT METAL 54 54 IRON 1 (BY SIMILARITY).  
FT METAL 58 58 IRON 1 AND 2 (BY SIMILARITY).  
FT METAL 73 73 IRON 2 (BY SIMILARITY).  
FT METAL 77 77 IRON 2 (BY SIMILARITY).  
FT METAL 101 101 IRON 2 (BY SIMILARITY).  
FT METAL 106 106 IRON 1 AND 2 (BY SIMILARITY).  
FT VARIANT 3 3 P -> E.  
FT VARIANT 10 10 W -> D.  
FT VARIANT 60 60 A -> G.  
FT VARIANT 66 66 K -> N.  
FT VARIANT 83 83 K -> Q.  
SQ SEQUENCE 113 AA; 12437 MW; 653C278D78F9E953 CRC64;  
Query Match 63.8%; Score 37; DB 1; Length 113;  
Best Local Similarity 70.0%; Pred. No. 9.2;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 1 AKFGGMGAKK 10  
Db 65 AKYGGYAHK 74  
RESULT 3  
DNAME\_CLOAB STANDARD; PRT; 374 AA.  
ID DNAME\_CLOAB P30725;  
AC 01-APR-1993 (Rel. 25, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
Chaperone protein dnaJ.  
DNAME\_CLOAB OR CAC1283.  
OC Clostridium acetobutylicum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;  
OC Clostridium.  
OX NCBI\_Taxid=1488;  
RN [1]  
RN SEQUENCE FROM N.A.  
RC STRAIN-ATCC 4259 / DSM 1731 / NCIB 619;  
RX MEDLINE=94123950; PubMed=7507453;  
RA Behrens S., Narberhaus F., Bahl H.;  
RT "Cloning, nucleotide sequence and structural analysis of the  
Clostridium acetobutylicum dnaJ gene";  
RL FEBS Microbiol. Lett. 114:53-60(1993).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC STRAIN-ATCC 824 / DSM 792 / VKM B-1787;  
RX MEDLINE=21359325; PubMed=11466286;  
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,  
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hiltl J., Wolf Y.I.,  
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,  
RA Bennett G.N., Koonin E.V., Smith D.R.;  
RT "Genome sequence and comparative analysis of the solvent-producing  
bacterium Clostridium acetobutylicum";  
RL J. Bacteriol. 183:4823-4838(2001).  
RN [3]

RP SEQUENCE OF 1-72 FROM N.A.  
RC STRAIN-ATCC 4259 / DSM 1731 / NCIB 619;  
RX MEDLINE=92250425; PubMed=1577695;  
RA Narberhaus F., Giebler K., Bahl H.;  
RT "Molecular characterization of the dnaJ gene region of Clostridium  
acetobutylicum, including gpe, dnaJ, and a new heat shock gene";  
RL J. Bacteriol. 174:3290-3299(1992).  
CC -1- FUNCTION: ACTS AS A CO-CHAPERONE. STIMULATES, JOINTLY WITH GRE,  
THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).  
CC -1- COFACTOR: BINDS TWO ZINC IONS PER MONOMER (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).  
CC -1- INDUCTION: BY HEAT SHOCK.  
CC -1- SIMILARITY: BELONGS TO THE DNAJ FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 J DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 CR DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X69050; CAA48792.1; -;  
DR EMBL: AE007640; AAK79254.1; -;  
DR EMBL: M74569; AAA23247.1; -;  
DR PIR: A41873; A41873.  
DR HSP; P25685; IHUJ.  
DR InterPro: IPR003095; DnaJ.  
DR InterPro: IPR002939; DnaJ\_C.  
DR InterPro: IPR001305; DnaJ\_CXXCXGXG.  
DR InterPro: IPR001623; DnaJ\_N.  
DR Pfam: PF00226; DnaJ; 1.  
DR Pfam: PF01556; DnaJ\_C; 1.  
DR Pfam: PF00684; DnaJ\_CXXCXGXG; 1.  
DR PRINTS: PR00625; DnaJPROTEIN.  
DR SMART: SM00271; DnaJ; 1.  
DR PROSITE: PS00636; DnaJ\_1; 1.  
DR PROSITE: PS00637; DnaJ\_2; 1.  
DR PROSITE: PS00637; DnaJ\_CXXCXGXG; 1.  
KW Chaperone; DNA replication; Heat shock; Repeat; Zinc; Metal-binding;  
KW Complete proteome.  
FT DOMAIN 5 70 J-DOMAIN.  
FT DOMAIN 76 108 GLY-RICH.  
FT REPEAT 150 157 CXXCXGXG MOTIF.  
FT REPEAT 167 174 CXXCXGXG MOTIF.  
FT REPEAT 193 200 CXXCXGXG MOTIF.  
FT REPEAT 207 214 CXXCXGXG MOTIF.  
FT METAL 150 150 ZINC 1 (BY SIMILARITY).  
FT METAL 153 153 ZINC 1 (BY SIMILARITY).  
FT METAL 167 167 ZINC 2 (BY SIMILARITY).  
FT METAL 170 170 ZINC 2 (BY SIMILARITY).  
FT METAL 193 193 ZINC 2 (BY SIMILARITY).  
FT METAL 196 196 ZINC 2 (BY SIMILARITY).  
FT METAL 207 207 ZINC 1 (BY SIMILARITY).  
FT METAL 210 210 ZINC 1 (BY SIMILARITY).  
SQ SEQUENCE 374 AA; 40401 MW; 0DC6B556511F201E CRC64;  
Query Match 63.8%; Score 37; DB 1; Length 374;  
Best Local Similarity 87.5%; Pred. No. 26;  
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 4 GGMGAKK 11  
Db 154 GGTGAKK 161  
RESULT 4  
RM62\_DROME STANDARD; PRT; 575 AA.  
ID RM62\_DROME P19109; Q9VNR4;  
DT 01-NOV-1990 (Rel. 16, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Putative ATP-dependent RNA helicase p62.  
 GN RM62 OR P62 OR CG10279.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91016833; PubMed=2170937;  
 RA Dorier D.R., Christensen A.C., Johnson D.H.;  
 RT "A novel RNA helicase gene tightly linked to the Triplo-lethal locus  
 RL Nucleic Acids Res. 18:5489-5494(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 PX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer G.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Boltslavsky S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimble B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laske P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclik J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Sytkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).  
 CC -1- SIMILARITY: BELONGS TO THE DEAD BOX HELICASE FAMILY. DDX5/DDX17  
 CC SUBFAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: X52846; CAA37037.1;  
 DR EMBL: AE003601; AAF51926.1; ALT\_INIT.  
 DR EMBL: AE003601; AAG22213.1; ALT\_INIT.

DR EMBL: AE003601; AAG22212.1; ALT\_INIT.  
 DR PIR: S11485; S11485.  
 DR HSSP: 058083; 1HV8.  
 DR FlyBase: FBgn0003261; Rm62.  
 DR InterPro: IPR001410; DEAD.  
 DR InterPro: IPR000629; DEAD\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.  
 DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; helicase\_C; 1.  
 DR SMART: SM00487; DEXDC; 1.  
 DR SMART: SM00490; HELIC; 1.  
 DR PROSITE: PS00039; DEAD\_ATP\_HELICASE; 1.  
 KW ATP-binding; RNA-binding; Helicase; Nuclear protein.  
 FT DOMAIN 9 81  
 FT NP\_BIND 181 188  
 FT SITE 291 298  
 FT COMPLET 527 569  
 FT CONFLICT 51 51  
 FT CONFLICT 531 531  
 SQ SEQUENCE 575 AA: 62474 MW: 77007C1A9DE51E9 CRC64;  
 Query Match 63.8%; Score 37; DB 1; Length 575;  
 Best Local Similarity 70.0%; Pred. No. 38;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 QY 2 KFGMGAKKG 11  
 Db 542 RFGGGFKKG 551  
 RESULT 5  
 ID DLP4\_HUMAN STANDARD; PRT; 989 AA.  
 AC Q9Y2H0; Q9H1L7; Q9H137; Q9H138;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Disks large-associated protein 4 (DAP-4) (SAP90/PSD-95-associated  
 DE protein 4) (SAPAP) (PSD-95/SAP90 binding protein 4).  
 DE DlgAP4 OR DAP4 OR KIA0964.  
 GN Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Brain;  
 RX MEDLINE=99246063; PubMed=10231032;  
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,  
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. XIII.  
 RT The complete sequences of 100 new cDNA clones from brain which code  
 RL for large proteins in vitro.";  
 RL DNA Res. 6:63-70(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Beasley D.V., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
 RA Buck D., Burrill W., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Grafham D.V., Griffiths S., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levanstain M.H., Leverisa M., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McConachie L.J., McLeay J., McMurray A.A.,  
 RA Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sena H.K., Showkeen R., Sims S.,  
 RA Skuce C.D., Smith M.L., Soderlund C., Steward A., Sulston J.E.,  
 RA Swann M., Symcox N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Way P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RT Nature 414:865-871(2001).  
 CC -1- FUNCTION: May play a role in the molecular organization of  
 CC synapses and neuronal cell signaling. Could be an adapter protein  
 CC linking ion channel to the subsynaptic cytoskeleton. May induce  
 CC enrichment of PSD-95/SAP90 at the plasma membrane.  
 CC -1- SUBUNIT: Interacts with DLG1 and DLG4/PSD-95 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; may be  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: BELONGS TO THE SAPAP FAMILY.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: AB023181; BAA76808.1; -  
 DR EMBL: AL390374; CAC18627.1; -  
 DR EMBL: AL050318; CAB75370.1; -  
 DR EMBL: AL050318; CAC16960.1; -  
 DR KW Membrane; Alternative splicing.  
 FT DOMAIN 267 274 POLY-PRO.  
 FT VARSPIC 671 697 EPTRRNGSLSEDNQPKADVAPSSSE -> VDCIQPPVKE  
 FT EPTSPATKFSIGVQVEDWR (IN ISOFORM 2).  
 FT CONFLICT 229 229 I -> T (IN REF. 2).  
 FT SEQUENCE 989 AA; 107578 MW; B391716801F6C1B CRC64;  
 SO  
 Query Match 63.8%; Score 37; DB 1; Length 989;  
 Best Local Similarity 70.0%; Pred. No. 62;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 Oy 2 KEGMGAKKG 11  
 DB 163 KVGNGSKKG 172  
 JULT 6  
 PAPA\_ECOLI STANDARD; PRT; 185 AA.  
 ID PAPA\_ECOLI P04127;  
 AC 01-NOV-1986 (Rel. 03, Created)  
 DT 01-NOV-1986 (Rel. 03, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PAP fibrillar major pilin protein precursor (PAP pilin).  
 GN PAPA.  
 OS Escherichia coli.  
 CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 CC Escherichia.  
 CC NCBI\_TaxID=562;  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=J96;  
 RX MEDLINE=84087728; PubMed=6140260;  
 RA Baga M., Normark S., Hardy J., O'Hanley P., Lark D., Olsson O.,  
 RA Schoonlik G., Falkow S.;  
 RT "Nucleotide sequence of the PAPA gene encoding the Pap pilus subunit  
 RT of human uropathogenic Escherichia coli.";  
 RL J. Bacteriol. 157:330-333(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=J96;

RX MEDLINE=93023852; PubMed=1357526;  
 RA Marling B.-I., Tennent J.M., Garcia E., Hamers A., Baga M.,  
 RA Lindberg F., Gastra W., Normark S.;  
 RT "Horizontal gene transfer of the Escherichia coli pap and pils pil  
 RT operators as a mechanism for the development of tissue-specific  
 RT adhesive properties.";  
 RL Mol. Microbiol. 6:2225-2242(1992).  
 CC -1- FUNCTION: FIMBRIN (ALSO CALLED PILI), POLAR FILAMENTS RADIATING  
 CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5  
 CC MICROMETERS AND NUMBERING 100-300 PER CELL. ENABLE BACTERIA TO  
 CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.  
 CC -1- SUBCELLULAR LOCATION: Secreted.  
 CC -1- DISEASE: STRAINS OF E. COLI THAT CAUSE INFECTION OF THE HUMAN  
 CC URINARY TRACT PRODUCE PAP-PILI WHICH ARE HAIR-LIKE APPENDAGES  
 CC CONSISTING OF ABOUT 1000 HELICALLY ARRANGED SUBUNITS OF THE  
 CC PROTEIN PAPA. THESE PILI MEDIANE BINDING TO DIGAUCOSIDE-  
 CC CONTAINING GLYCOPOLIDS PRESENT ON THE EPITHELIAL CELLS WHICH  
 CC LINE THE URINARY TRACT.  
 CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIN PROTEINS.  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: X03391; CAA27126.1; -  
 DR EMBL: X61239; CAA43562.1; -  
 DR PIR: A23221; YOECP.  
 DR InterPro: IPR000259; Fimbrin.  
 DR Pfam: PF00419; Fimbrin; 1.  
 DR KW Fimbria; Signal.  
 FT SIGNAL 1 22  
 FT CHAIN 23 185 PAP FIMBRIAL MAJOR PILIN PROTEIN.  
 FT DISULFID 44 83 PROBABLE.  
 FT SEQUENCE 185 AA; 18686 MW; 93DB4FFDA211C671 CRC64;  
 SO  
 Query Match 62.1%; Score 36; DB 1; Length 185;  
 Best Local Similarity 87.5%; Pred. No. 21;  
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 Oy 4 GGMGAKKG 11  
 DB 90 GGMGAKKG 97  
 RESULT 7  
 MEHS\_WOLSU STANDARD; PRT; 354 AA.  
 ID MEHS\_WOLSU P31884;  
 AC 01-JUL-1993 (Rel. 26, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 01-FEB-1996 (Rel. 33, Last annotation update)  
 DE Quinone-reactive Ni/Fe-hydrogenase small chain precursor  
 DE (EC 1.12.99.3) (Membrane-bound hydrogenase small subunit)  
 DE (Hydrogen:quinone oxidoreductase).  
 GN HYDA.  
 OS Wolinella succinogenes.  
 CC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;  
 CC Wolinella.  
 CC NCBI\_TaxID=844;  
 OX NCBI\_TaxID=844;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 37-52.  
 RX MEDLINE=92267032; PubMed=158728;  
 RA Fahrenholz F., Kojro E., Duchene A., Tripier D., Juvenal K.,  
 RA Krogger A.;  
 RT "The quinone-reactive Ni/Fe-hydrogenase of Wolinella succinogenes.";  
 RL Eur. J. Biochem. 206:93-102(1992).  
 RN [2]  
 RP ERRATUM.

CC MEDLINE=93307313; PubMed=8319698;  
RA Dross F., Geisler V., Lenger R., Theis F., Krafft T.,  
RA Fahrenholz F., Kojro E., Duchene A., Tripier D., Juvenal K.,  
RA Kroeger A.;  
RL Eur. J. Biochem. 214:949-950(1993).  
CC -1- CATALYTIC ACTIVITY: H(2) + menaquinone = reduced menaquinone.  
CC -1- COFACTOR: CONTAINS NICKEL AND IRON-SULFUR.  
CC -1- SUBUNIT: HETERODIMER OF A LARGE AND A SMALL SUBUNIT.  
CC -1- SUBCELLULAR LOCATION: Membrane-bound.  
CC -1- SIMILARITY: TO OTHER UPTAKE HYDROGENASES SMALL SUBUNIT.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X65189; CAA46302.1; ALT\_INIT.  
DR PIR: S22404; S22404.  
DR HSSP: P21853; 1H2A.  
DR InterPro: IPR002096; Complex1\_20KD.  
DR InterPro: IPR001821; Nlfehydrog\_small.  
DR Pfam: PF01058; oxidored\_g6; 1.  
DR PRINTS: PR00614; NlfeHASESMUL.  
KW Oxidoreductase; Signal; Membrane; Iron-sulfur; Nickel.  
FT SIGNAL 1 36  
FT CHAIN 37 354  
FT CHAIN QUINONE-REACTIVE NI/FE-HYDROGENASE SMALL  
SQ SEQUENCE 354 AA; 38293 MW; C39CC21D2F5A87D0 CRC64;  
Query Match 62.1%; Score 36; DB 1; Length 354;  
Best Local Similarity 70.0%; Pred. No. 37;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 2 KFGMGAKKG 11  
DB 241 EFGDEGAKKG 250  
ID DLP4\_RAT  
AC P97839;  
STANDARD: PRT; 992 AA.  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DT Disks large-associated protein 4 (DAP-4) (SAP90/PSD-95-associated  
protein 4) (SAPAP4) (PSD-95/SAP90 binding protein 4).  
DI DAP4 OR DAP4.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=97277335; PubMed=9115257;  
RA Takachul M., Hata Y., Hirao K., Toyoda A., Irie M., Takai Y.;  
RA "SAPAPs, A family of PSD-95/SAP90-associated proteins localized at  
RT postsynaptic density.";  
RL J. Biol. Chem. 272:11943-11951(1997).  
CC -1- FUNCTION: May play a role in the molecular organization of  
CC synapses and neuronal cell signaling. Could be an adapter protein  
CC linking ion channel to the subynaptic cytoskeleton. May induce  
CC enrichment of PSD-95/SAP90 at the plasma membrane.  
CC -1- SUBUNIT: Interacts with DLG1 and DLG4/PSD-95 (By similarity).  
CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).  
CC -1- TISSUE SPECIFICITY: Expressed in brain.  
CC -1- SIMILARITY: BELONGS TO THE SAPAP FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: U67140; AAB48590.1; -.  
KW Membrane.  
FT DOMAIN 266 273  
FT DOMAIN POLY-PRO.  
SQ SEQUENCE 992 AA; 108034 MW; 8875B67C3FD71F04 CRC64;  
Query Match 62.1%; Score 36; DB 1; Length 992;  
Best Local Similarity 70.0%; Pred. No. 92;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 KFGMGAKKG 11  
DB 163 KVGNGGKKG 172  
ID RS19\_MYCCA  
AC P10132;  
STANDARD: PRT; 88 AA.  
DT 01-MAR-1989 (Rel. 10, Created)  
DT 01-MAR-1989 (Rel. 10, Last sequence update)  
DT 01-OCT-1994 (Rel. 30, Last annotation update)  
DE 30S ribosomal protein S19.  
GN RPS5.  
OS Mycoplasma capricolum.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollitutes;  
OC Entomoplasmataceae.  
OX NCBI\_Taxid=2095;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 27343 / K12;  
RX MEDLINE=88142549; PubMed=3481422;  
RA Okubo S., Muto A., Kawachi Y., Yamao F., Osawa S.;  
RA "The ribosomal protein gene cluster of Mycoplasma capricolum.";  
RL Mol. Genet. 210:314-322(1987).  
CC -1- FUNCTION: PROTEIN S19 FORMS A COMPLEX WITH S13 THAT BINDS STRONGLY  
CC TO THE 16S RIBOSOMAL RNA (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE S19P FAMILY OF RIBOSOMAL PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X06414; CAA29708.1; -.  
DR PIR: S02835; R3YM19.  
DR HSSP: P80381; 1OKF.  
DR InterPro: IPR002222; Ribosomal\_S19.  
DR Pfam: PF00203; Ribosomal\_S19; 1.  
DR PRINTS: PR00975; RIBOSOMAL\_S19.  
DR ProDom: PD001012; Ribosomal\_S19; 1.  
DR PROSITE: PS00323; RIBOSOMAL\_S19; 1.  
KW Ribosomal protein; rRNA-binding.  
SQ SEQUENCE 88 AA; 9903 MW; 1BF2AC24D697235E CRC64;  
Query Match 60.3%; Score 35; DB 1; Length 88;  
Best Local Similarity 70.0%; Pred. No. 15;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 KFGMGAKKG 11  
DB 76 KFGHDDKG 85

```

RESULT 10
RL3_PLARO STANDARD; PRT; 140 AA.
ID AH11_YEAST
AC P72233;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S ribosomal protein L3 (Fragment).
GN RPLC.
OS Planobispora rosea.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptosporangineae; Streptosporangiaceae;
OC Planobispora.
OX NCBI_TaxID=35762;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 53733;
RX MEDLINE=97055420; PubMed=8899707;
PA Sosis M., Amati G., Cappellano C., Sarubbi E., Monti F.,
PA Donadio S.;
PA "An elongation factor Tu (EF-Tu) resistant to the EF-Tu inhibitor
PA GE2270 in the producing organism Planobispora rosea.";
RL Mol. Microbiol. 22:43-51(1996).
CC -1- FUNCTION: THIS PROTEIN BINDS DIRECTLY TO 23S RIBOSOMAL RNA AND MAY
CC PARTICIPATE IN THE FORMATION OF THE PEPTIDYLTRANSFERASE CENTER OF
CC THE RIBOSOME (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE L3P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X98830; CAA67347.1; -.
CC InterPro: IPR000597; Ribosomal_L3.
CC Pfam: PF00297; Ribosomal_L3; 1.
CC DR PRODOM: PD001374; Ribosomal_L3; 1.
CC DR PROSITE: PS00474; RIBOSOMAL_L3; 1.
CC KM Ribosomal protein; rRNA-binding.
CC FT NON_TER 140 140
CC SQ SEQUENCE 140 AA; 15232 MW; B50630F7A0A4849D CRC64;

Query Match 60.3%; Score 35; DB 1; Length 140;
Best Local Similarity 66.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

3 FGGMGARKG 11
111:111
Db 131 FGGGASHG 139

RESULT 11
AH11_YEAST STANDARD; PRT; 182 AA.
ID AH11_YEAST
AC P29589; P38807;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE Hexose transport activator protein.
GN AH11 OR YHR093W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC 971 B;
RA Gezcan S., Cliracy M.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
[2]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=94378003; PubMed=8091229;
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA Kucaba T., Hillier L., Jier M., Johnston L., Langston Y.,
RA Latreille P., Louis E.J., Macri C., Mardis E., Meneses S., Mouser L.,
RA Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA Vignati D., Wilcox L., Wohlman P., Waterston R., Wilson R.,
RA Vaudin M.;
RT "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT VIII.";
RL Science 265:2077-2082(1994).
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X59464; CAA42071.1; -.
CC DR EMBL: U00060; AAB68925.1; -.
CC DR PIR: S17005; S17005.
CC DR PIR: S46717; S46717.
CC DR SGD: S0001135; AH11.
CC SQ SEQUENCE 182 AA; 19776 MW; 7B0723BAA1066713 CRC64;

Query Match 60.3%; Score 35; DB 1; Length 182;
Best Local Similarity 70.0%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

2 KFGMGARKG 11
111:111
Db 54 KFGGAKKG 63

RESULT 12
RA55_SCHPO STANDARD; PRT; 350 AA.
ID RA55_SCHPO
AC O14128;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA repair protein rhp55 (RAD55 homolog).
GN RHP55 OR SPAC3C7.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA Bashkirtov V.I., Khasanov F.K., Savchenko G.V., Heyer W.-D.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Brown D., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RL -1- SUBCELLULAR LOCATION: Nuclear (potential).
CC -1- SIMILARITY: STRONG TO OTHER EUKARYOTIC RECA-LIKE PROTEIN; SOME, TO
CC PROKARYOTIC RECA PROTEIN.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----

```



DR EMBL: AF053410; AAC17871.1; -  
 DR EMBL: Z99568; CAB16734.1; -  
 DR InterPro: IPR001553; RECA.  
 DR PRINTS: PR00142; RECA.  
 DR PROSITE: PS50162; RECA\_2; 1.  
 KW DNA damage; DNA repair; ATP-binding; Nuclear protein.  
 FT SMP BIND 51 58 ATP (POTENTIAL).  
 FT SEQUENCE 350 AA; 38996 MW; 53DC65C0EC3836E1 CRC64;

Query Match 60.3%; Score 35; DB 1; Length 350;  
 Best Local Similarity 66.7%; Pred. NO. 55;  
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKKG 11  
 ||| |||  
 DB 36 FCGSGLRKG 44

RESULT 13  
 CH6B\_DROME STANDARD; PRT; 648 AA.  
 Q9VPS5:  
 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Probable 60 kDa heat shock protein homolog 1, mitochondrial precursor  
 DE (Hsp60) (60 kDa chaperonin) (CPN60) (Heat shock protein 60) (HSP-60).  
 GN HSP60B OR CG2830.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 [1]  
 SEQUENCE FROM N.A.  
 RP STRAIN=BERKELEY;  
 RC MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Ffiankoch C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibbegan C.,  
 RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kilp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shee B.C., Siden-Kiamos I., Simpson M.C., Skupski M.P., Smith T.,  
 RA Splyer E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Styksas R., Tector C., Turner C., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 CC -I- FUNCTION: PREVENTS MISFOLDING AND PROMOTES THE REFOOLDING AND  
 CC PROPER ASSEMBLY OF UNFOLDED POLYPEPTIDES GENERATED UNDER STRESS  
 CC CONDITIONS (BY SIMILARITY).  
 CC -I- SUBCELLULAR LOCATION: Mitochondrial matrix (By similarity).  
 CC -I- SIMILARITY: BELONGS TO THE CHAPERONIN (HSP60) FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 CC EMBL: AE003588; AAF51467.1; -  
 DR HSPB; P06139; 1GRL.  
 DR FLYBASE; FBgn0011244; Hsp60B.  
 DR InterPro: IPR001844; Chaperonins\_cpn60.  
 DR InterPro: IPR002423; TCPL\_cpn60.  
 DR Pfam: PF00118; cpn60\_TCPL\_1.  
 DR PRINTS: PR00298; CHAPERONIN60.  
 DR PRINTS: PR00304; TCOMPLEXTCPL.  
 DR PROSITE: PS00296; CHAPERONINS\_CPN60; 1.  
 KW Hypothetical protein; Chaperone; ATP-binding; Mitochondrion;  
 KW Transit peptide.  
 FT TRANSIT 1 55 MITOCHONDRION (BY SIMILARITY).  
 FT CHAIN 56 648 PROBABLE 60 KDA HEAT SHOCK PROTEIN  
 FT HOMOLOG 1.  
 SQ SEQUENCE 648 AA; 68636 MW; E45D4A4166C78455 CRC64;

Query Match 60.3%; Score 35; DB 1; Length 648;  
 Best Local Similarity 77.8%; Pred. NO. 94;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKKG 11  
 ||||| |  
 DB 602 FCGMGAGGG 610

RESULT 14  
 NCNG4\_HUMAN STANDARD; PRT; 909 AA.  
 ID NCNG4\_HUMAN  
 AC Q14028; Q14029;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Cyclic-nucleotide-gated cation channel 4 (CNG channel 4) (CNG-4)  
 DE (CNG4) (Cyclic nucleotide-gated cation channel modulatory subunit).  
 GN CNGB1 OR CNGC4.  
 OS Homo sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Euteria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 SEQUENCE FROM N.A.  
 RP TISSUE=Retina;  
 RC MEDLINE=93226050; PubMed=7682292;  
 RA Chen T.Y., Peng Y.-W., Dhallan R.S., Ahmed B., Reed R.R., Yau K.-W.;  
 RA "A new subunit of the cyclic nucleotide-gated cation channel in  
 RA retinal rods.";  
 RL Nature 362:764-767(1993).  
 RL -I- SUBUNIT: HETEROOLIGOMERIC COMPLEX WITH CNG1.  
 CC -I- SUBCELLULAR LOCATION: Integral membrane protein.  
 CC -I- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RCNC2A AND RCNC2B (SHOWN  
 CC HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -I- SIMILARITY: BELONGS TO THE CYCLIC NUCLEOTIDE-GATED CATION CHANNEL  
 CC FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, L15296; AAA65620.1; -;  
DR EMBL, L15297; AAA65619.1; -;  
DR MIM; 600724; -;  
DR InterPro: IPR000636; Catlon\_chan\_non\_11g.  
DR InterPro: IPR000595; CNMP\_binding.  
DR Pfam: PF000627; CNMP\_binding; 1.  
DR Pfam: PF00520; Ion\_trans; 1.  
DR SMART; SM00100; CNMP; 1.  
DR PROSITE; PS00888; CNMP\_BINDING\_1; 1.  
DR PROSITE; PS00889; CNMP\_BINDING\_2; 1.  
DR PROSITE; PS50042; CNMP\_BINDING\_3; 1.  
KW Ionic channel; Ion transport; CAMP-binding; Transmembrane;  
KW Multigene family; Alternative splicing.  
FT DOMAIN 1 314 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 315 333 H1 (POTENTIAL).  
FT DOMAIN 334 347 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 348 366 H2 (POTENTIAL).  
FT DOMAIN 368 391 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 392 411 H3 (POTENTIAL).  
FT DOMAIN 412 448 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 449 471 H4 (POTENTIAL).  
FT DOMAIN 472 515 CYTOPLASMIC (POTENTIAL).  
FT TRANSSEM 516 535 H5 (POTENTIAL).  
FT DOMAIN 536 619 EXTRACELLULAR (POTENTIAL).  
FT TRANSSEM 620 640 H6 (POTENTIAL).  
FT DOMAIN 641 909 CYTOPLASMIC (POTENTIAL).  
FT NE\_BIND 9 29 POLY-GLU.  
FT BINDING 628 767 CAMP (BY SIMILARITY).  
FT BINDING 688 700 CAMP (POTENTIAL).  
FT VARSPLIC 700 700 CAMP (POTENTIAL).  
SQ SEQUENCE 909 AA: 102285 MW: DCOE75436B6EDD CRC64;  
MISSING (IN ISOFORM RCNC2A).  
Query Match 60.3%; Score 35; DB 1; Length 909;  
Best Local Similarity 70.0%; Pred. No. 1.3e+02;  
Matches 7; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
QY 2 FGGMGAKKG 11  
DB 776 KMGKGAKKG 785  
-----  
TITLE 15  
ID YEAST  
AB Y1A2\_YEAST STANDARD; PRT; 946 AA.  
AC P40559;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Hypothetical 108.4 kDa protein in BERY-PANI intergenic region.  
GN Y1L002C OR Y1A2C.  
OS Saccharomyces cerevisiae (Baker's yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
OX NCBI\_TaxID=4932;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C / AB972;  
RA Barrett B.G., Badcock K., Bankier A.T., Bowman S., Brown D.,  
RA Churcher C.M., Connor R., Copsey T., Dear S., Devlin K., Fraser A.,  
RA Gentles S., Hamlyn N., Hornslell T.S., Hunt S., Jagers K., Jones M.,  
RA Louis E., Lye G., Moule S., Moule T., Odell C., Pearson D.,  
RA Rajandream M.A., Riles L., Rowley N., Skelton J., Smith V.,  
RA Walsh S.V., Whitehead S.;  
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=5288C;

FX MEDLINE=95282515; PubMed=7762303;  
RA Voss H., Tamames J., Teodoru C., Valencia A., Sensen C., Wiemann S.,  
RA Schwager C., Zimmermann J., Sander C., Ansoorge W.;  
RT "Nucleotide sequence and analysis of the centromeric region of yeast  
RT chromosome IX.";  
RL Yeast 11:61-78(1995).  
CC -1- SIMILARITY: BELONGS TO THE INOSITOL-1,4,5-TRISPHOSPHATE  
CC -1- PHOSPHATASE TYPE II FAMILY.  
CC -1- SIMILARITY: CONTAINS 1 SAC DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL, Z38062; CAA86201.1; -;  
DR EMBL, X79743; -; NOT\_ANNOTATED\_CDS.  
DR PIR; S48433; S48433.  
DR SGD; S0001264; Y1L002C.  
DR InterPro: IPR000300; IPIC.  
DR InterPro: IPR002013; SYJA\_N.  
DR Pfam; PF00783; IPIC; 1.  
DR Pfam; PF02383; SYJA\_N; 1.  
DR SMART; SM00128; IPIC; 1.  
DR PROSITE; PS50275; SAC; 1.  
KW Hypothetical protein; Hydrolase.  
FT DOMAIN 151 480 SAC.  
SQ SEQUENCE 946 AA: 108429 MW: A833C39B0A62543F CRC64;

Query Match 60.3%; Score 35; DB 1; Length 946;  
Best Local Similarity 66.7%; Pred. No. 1.3e+02;  
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 3 FGGMGAKKG 11  
DB 644 FGGMASNKG 652

Search completed: October 28, 2002, 17:21:41  
Job time: 10.1053 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 : Search time 25.4737 Seconds

(without alignments)  
74.702 Million cell updates/sec

Title: US-09-833-079-2  
Perfect score: 58  
Sequence: 1 AKFGMGAKKG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
al number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: sp.archaea:\*  
2: sp.bacteria:\*  
3: sp.fungi:\*  
4: sp.human:\*  
5: sp.invertebrate:\*  
6: sp.mammal:\*  
7: sp.mhc:\*  
8: sp.organelle:\*  
9: sp.phage:\*  
10: sp.plant:\*  
11: sp rodent:\*  
12: sp.virus:\*  
13: sp.vertebrate:\*  
14: sp.unclassified:\*  
15: sp.virus:\*  
16: sp.bacteriophage:\*  
17: sp.archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	67.2	237	16	0915A7
2	39	67.2	1173	5	095XK7
3	38	65.5	126	5	044355
4	38	65.5	132	5	044357
5	38	65.5	136	5	044354
6	38	65.5	156	10	095XK7
7	38	65.5	516	3	090UB5
8	38	65.5	558	3	096TU2
9	37	63.8	142	16	092F57
10	37	63.8	371	17	097W08
11	37	63.8	578	5	0917P5
12	37	63.8	629	13	090WR5
13	36	62.1	131	5	044356
14	36	62.1	366	10	094DD1
15	36	62.1	387	17	058754
16	36	62.1	399	17	09YEF2

17	36	62.1	485	2	09L656	091656 streptomyc
18	36	62.1	485	16	098MA1	098ma1 rhizobium 1
19	36	62.1	608	2	09AJC4	09ajc4 chlorobium
20	36	62.1	886	3	09VGC3	09vgc3 drosophila
21	36	62.1	1292	3	096WLO	096wlo ustilago ma
22	36	62.1	1814	5	09BLM9	09blm9 toxocara ca
23	36	62.1	1957	5	004009	004009 brugia mala
24	36	62.1	1957	5	004010	004010 onchocerca
25	35	60.3	113	5	09V5U5	09v5u5 drosophila
26	35	60.3	235	10	09EF85	09ef85 arabidopsis
27	35	60.3	338	15	090DY4	090dy4 porcine end
28	35	60.3	292	16	09ABX3	09abx3 caulobacter
29	35	60.3	305	17	097VA3	097va3 sulfolobus
30	35	60.3	316	16	010881	010881 mycobacteri
31	35	60.3	321	1	09V2U2	09v2u2 methanococ
32	35	60.3	395	17	09HL40	09hl40 thermoplas
33	35	60.3	579	3	09US10	09us10 schizosacch
34	35	60.3	604	1	093700	093700 sulfolobus
35	35	60.3	639	5	0961V0	0961v0 drosophila
36	35	60.3	654	15	090RL8	090rl8 porcine end
37	35	60.3	660	15	041172	041172 porcine end
38	35	60.3	661	15	0910F6	091uf6 porcine end
39	35	60.3	721	3	09UT74	09ut74 schizosacch
40	35	60.3	753	2	087472	087472 treponema p
41	35	60.3	756	16	083337	083337 treponema p
42	35	60.3	758	2	09KHF2	09khf2 treponema p
43	35	60.3	758	2	09KHF1	09khf1 treponema p
44	35	60.3	758	2	09KHF0	09khf0 treponema p
45	35	60.3	758	16	007894	007894 treponema p

## ALIGNMENTS

## RESULT 1

0915A7 ID 0915A7 PRELIMINARY: PRT: 237 AA.

AC 0915A7; 01-MAR-2001 (TREMBLrel, 16, Created)

DT 01-MAR-2001 (TREMBLrel, 16, Last sequence update)

DT 01-OCT-2001 (TREMBLrel, 18, Last annotation update)

DE HYPOTHETICAL PROTEIN PA0833.

GN PA0833.

OS Pseudomonas aeruginosa.

OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;

OC Pseudomonas.

OX NCBI\_Taxid=287;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC 15692 / PA01;

RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,

RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,

RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,

RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbis K., Lim R.M.,

RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,

RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;

RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";

RL Nature 406:959-964(2000).

RL EMBL, AB004518; AAC04222.1; -

DR InterPro: IPR001145; Bac\_OmpA.

DR Pfam: PF00691; OmpA; 1.

DR PRINTS: PR01021; OMPADOMAIN.

DR PRODOM: PD000930; Bac\_OmpA; 1.

KW Hypothetical protein: Complete proteome.

SO SEQUENCE 237 AA; 24713 MW; 4DB8BA8ABDC8F00 CRC64;

Query Match 67.2%; Score 39; DB 16; Length 237;

Best Local Similarity 63.6%; Pred. No. 28;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKFGMGAKKG 11

Db 42 AKYGGALAG 52

## RESULT 2

095XK7 PRELIMINARY; PRT; 1173 AA.

ID 095XK7

AC 095XK7

DT 01-DEC-2001 (TRENBLREL. 19, Created)

DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)

DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)

DE HYPOTHEICAL 133.3 KDA PROTEIN.

GN Y54F10BM.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Pelodierinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA MEDLINE=99069613; PubMed=9851916;

RT None;

RT "Genome sequence of the nematode C. elegans: a platform for

investigating biology. The C. elegans Sequencing Consortium.";

RL Science 282:2012-2018(1998).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Bradshaw-Cordum H., Ryan E., Courtney L., Yeakum M.;

RT "The sequence of C. elegans cosmid Y54F10BM.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA Waterston R.;

RT "Direct Submission.";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC026301; AAK68893.1; -.

KW Hypothetical protein.

SQ SEQUENCE 1173 AA; 133251 MW; 1E5D84ABF21A725 CRC64;

Query Match 67.2%; Score 39; DB 5; Length 1173;

Best Local Similarity 80.0%; Pred. NO. 1.5e+02;

Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 FGGMGAKKG 11

DB 205 KKGAGAKKG 214

RESULT 3

044355 PRELIMINARY; PRT; 126 AA.

ID 044355

AC 044355

DT 01-JUN-1998 (TRENBLREL. 06, Created)

DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)

DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)

DE ABDUCTIN.

OS Aequipecten irradians (Bay scallop).

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;

OC Pectinoidea; Pectinidae; Argopecten.

OX NCBI\_TaxID=31199;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98044338; PubMed=9382816;

RA Cao Q.-P., Wang Y.-J., Bayley H.;

RT "Sequence of abductin, the molluscan 'rubber' protein.";

RL Curr. Biol. 7:0-0(1997).

DR EMBL; AF026846; AAB94678.1; -.

SQ SEQUENCE 126 AA; 10593 MW; CFEED657E9CC3369 CRC64;

Query Match 65.5%; Score 38; DB 5; Length 126;

Best Local Similarity 77.8%; Pred. No. 21;

QY 3 FGGMGAKKG 11

DB 82 FGGMGGKG 90

RESULT 6

09SEX7 PRELIMINARY; PRT; 156 AA.

ID 09SEX7

AC 09SEX7

DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 FGGMGAKKG 11

DB 82 FGGMGGKG 90

## RESULT 4

044357 PRELIMINARY; PRT; 132 AA.

ID 044357

AC 044357

DT 01-JUN-1998 (TRENBLREL. 06, Created)

DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)

DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)

DE ABDUCTIN.

OS Aequipecten irradians (Bay scallop).

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;

OC Pectinoidea; Pectinidae; Argopecten.

OX NCBI\_TaxID=31199;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98044338; PubMed=9382816;

RA Cao Q.-P., Wang Y.-J., Bayley H.;

RT "Sequence of abductin, the molluscan 'rubber' protein.";

RL Curr. Biol. 7:0-0(1997).

DR EMBL; AF026848; AAB94680.1; -.

SQ SEQUENCE 132 AA; 11103 MW; 992561B8B5F4A608 CRC64;

Query Match 65.5%; Score 38; DB 5; Length 132;

Best Local Similarity 77.8%; Pred. NO. 23;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 FGGMGAKKG 11

DB 82 FGGMGGKG 90

## RESULT 5

044354 PRELIMINARY; PRT; 136 AA.

ID 044354

AC 044354

DT 01-JUN-1998 (TRENBLREL. 06, Created)

DT 01-JUN-1998 (TRENBLREL. 06, Last sequence update)

DT 01-NOV-1998 (TRENBLREL. 08, Last annotation update)

DE ABDUCTIN.

OS Aequipecten irradians (Bay scallop).

OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorpha; Pectinoidea;

OC Pectinoidea; Pectinidae; Argopecten.

OX NCBI\_TaxID=31199;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=98044338; PubMed=9382816;

RA Cao Q.-P., Wang Y.-J., Bayley H.;

RT "Sequence of abductin, the molluscan 'rubber' protein.";

RL Curr. Biol. 7:0-0(1997).

DR EMBL; AF026845; AAB94677.1; -.

SQ SEQUENCE 136 AA; 11365 MW; 11634DF07F935EC7 CRC64;

Query Match 65.5%; Score 38; DB 5; Length 136;

Best Local Similarity 77.8%; Pred. NO. 23;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 FGGMGAKKG 11

DB 82 FGGMGGKG 90

## RESULT 6

09SEX7 PRELIMINARY; PRT; 156 AA.

ID 09SEX7

AC 09SEX7

DT 01-MAY-2000 (TRENBLREL. 13, Created)

DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)

```

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)
PR PUTATIVE BZIP TRANSCRIPTION FACTOR.
GN AT2G17770.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Ronnaley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feildlyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.U., Ronning C.M., Koo H., Morfat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umeyam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Cressy T.H., Goodman H.M., Somerville C.R.,
RA Copenhagen G.P., Preuss D., Niernan W.C., White O., Eisen J.A.,
RA Salbeger S.L., Fraser C.M., Venter J.C.;
PT *Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.;
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RL Lin X.;
RM Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
CC CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL: AF024504; AAF18682.1; -.
DR InterPro: IPR001871; bzip.
DR Pfam: PF00170; bzip. 1.
DR PROSITE: PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 156 AA; 17476 MW; EA2F76C2F068922 CRC64;

Query Match 65.5%; Score 38; DB 10; Length 156;
Best Local Similarity 63.6%; Pred. No. 27;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKFGGAKRG 11
|||||:|:|:|
Db 106 AKFGCLGKRG 116

RESULT 7
Q9UTB5 PRELIMINARY; PRT; 516 AA.
Q9UTB5;
01-MAY-2000 (TREMBLrel. 13, Created)
01-MAY-2000 (TREMBLrel. 13, Last sequence update)
01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUTATIVE PHOSPHATIDYLKERINE DECARBOXYLASE PROENZYME 1.
SPAC2558.03.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Barrett B.G., Rajandream M.A., McDougall R.C., McLean J., Harris D.,
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133225; CAB61769.1; -.
DR InterPro: IPR003817; PS_Dcarboxylase; 1.
DR Pfam: PF02666; PS_Dcarboxylase; 1.
SQ SEQUENCE 516 AA; 58464 MW; FFE457955895FB9 CRC64;

Query Match 65.5%; Score 38; DB 3; Length 516;
Best Local Similarity 77.8%; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY	3	EGGCGAKKG 11	65.5%; Score 38; DB 3; Length 558;
	11111111		Pred. NO. 1.1e+02;
Db	22	EGGCGALKG 30	Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
RESULT 8			
096TU2	096TU2	PRELIMINARY; PRT; 558 AA.	
AC	096TU2;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
GN	CAP64 PROTEIN.		
GN	CAP64.		
OS	Pleurotus ostreatus (Oyster mushroom) (White-rot fungus).		
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;		
OC	Agaricales; Pleurotaceae; Pleurotus.		
OX	NCBI_TaxID=5322;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=001; TISSUE=LAMELLAE;		
RA	Valenzuela M., Park S.K., Penas M.M., Anaut M., Ramirez L.,		
RA	Pisabarro A.G.;		
RT	"Isolation and characterization of a gene of the edible basidiomycete		
RT	Pleurotus ostreatus that is homologous to CAP64 from Cryptococcus		
RT	neoformans.";		
RL	Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: AJ318523; CAC50236.1; -.		
SD	SEQUENCE 558 AA; 59934 MW; 226DAR03ELB78F8D CRC64;		
Query Match			
Best Local Similarity	65.5%; Score 38; DB 3; Length 558;		
Matches	6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;		
QY	3	EGGCGAKKG 11	
Db	347	EGGCGAKKG 355	
RESULT 9			
ID	092F57	PRELIMINARY; PRT; 142 AA.	
AC	092F57;		
DT	01-DEC-2001 (TREMBLrel. 19, Created)		
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)		
DE	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	POLYRIBONUCLEOTIDE NUCLEOTIDYLTRANSFERASE DOMAIN PRESENT.		
GN	LIN0250.		
OS	Listeria innocua.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
OC	Bacillus/Staphylococcus group; Listeria.		
OX	NCBI_TaxID=1642;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CLIP 11262 / SEROVAR 66;		
RX	PubMed=11679669;		
RA	Glaser P., Frangeul L., Buchrieser C., Rusnok C., Amend A.,		
RA	Baquerot F., Berche P., Bloeker H., Brandt P., Chakraborty T.,		
RA	Charbit A., Chetoui F., Couve E., de Daruvar A., Deloux P.,		
RA	Domian E., Dominguez-Bernal G., Duchud E., Durant L., Dussutget O.,		
RA	Entian K.-D., Fshli H., Garcia-del Portillo F., Garrido P.,		
RA	Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,		
RA	Jautier L.-M., Kaest U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,		
RA	Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari R.,		
RA	Nordstleik G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,		
RA	Remmel B., Rose M., Schueter T., Simoes N., Tierrez A.,		
RA	Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;		
RT	"Comparative genomics of Listeria species.";		
RL	Science 294:849-852(2001).		
DR	EMBL: AL596164; CAC95483.1; -.		
DR	Listlist: LIN00250; -.		
TX	Transferase; Complete proteome.		
SD	SEQUENCE 142 AA; 15647 MW; 7E4F086350AC3A9 CRC64;		

Query Match 63.8%; Score 37; DB 16; Length 142;  
 Best Local Similarity 63.6%; Pred. No. 37;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KFGGGAAGG 11  
 DB 132 SKRGGRGAKG 142

## RESULT 10

O97W08 PRELIMINARY; PRT; 371 AA.

AC 097W08; PRELIMINARY; PRT; 371 AA.  
 DT 01-OCT-2001 (TREMBlrel. 18, Created)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)  
 ACYL-COA DEHYDROGENASE (ACD-3) (EC 1.3.99.).

ACD-3.

Sulfolobus solfataricus.

Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.

NCBI\_TaxID=2287;

[1] SEQUENCE FROM N.A.

RC STRAIN-ATCC 35092 / DSM 1617 / P2;  
 MEDLINE-2133296; PubMed-11427726;

RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,  
 RA Aweyer M.J., Chau-Welher C.C.-Y., Clausen I.G., Curtis B.A.,  
 RA De Moors A., Eranos G., Fletcher C., Gordon P.M.K.,  
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,  
 RA Thi Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,  
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,  
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,  
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus p2.";  
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).

RL EMBL: AE006813; AKK42244.1; -  
 DR InterPro: IPR001552; Acyl-CoA\_dh.

DR Pfam: PF00441; Acyl-CoA\_dh; 1.  
 DR Pfam: PF02770; Acyl-CoA\_dh.M; 1.  
 DR Pfam: PF02771; Acyl-CoA\_dh.N; 1.

KW Oxidoreductase; Complete proteome.

SO SEQUENCE 371 AA; 41394 MW; A1B2C43E8A106884 CRC64;

Query Match 63.8%; Score 37; DB 17; Length 371;  
 Best Local Similarity 77.8%; Pred. No. 1e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FGGGGAAGG 11  
 DB 324 FGGGGAAGG 332

## RESULT 11

O917P5 PRELIMINARY; PRT; 578 AA.

AC 0917P5; PRELIMINARY; PRT; 578 AA.  
 DT 01-MAR-2001 (TREMBlrel. 16, Created)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE RM62 PROTEIN.

GN RM62 OR CG10279.

OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephyroidea; Drosophilidae; Drosophila.

NCBI\_TaxID=7227;

[1] SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;  
 MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
 RA Abrial J.F., Aspray A., An H.-J., Andrews-Plamkosh C., Baldwin D.,  
 RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalili M., Kalush F., Kapran G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matrel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mitsuhashi N.V., Mobarry C., Morris J., Mostrel A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E.C., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodard J., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).

CC -1- SIMILARITY: BELONGS TO THE "DEAD" BOX FAMILY HELICASE.

DR EMBL: AE003601; AAF51927.2; -  
 DR HSSP: Q58083; 1HV8.

DR FLYBase: FBgn0003261; Rm62.  
 DR InterPro: IPR001410; DEAD.

DR InterPro: IPR000629; DEAD\_ATP\_helicase.  
 DR InterPro: IPR001650; Helicase\_C.

DR InterPro: IPR000504; RRM.

DR Pfam: PF00270; DEAD; 1.  
 DR Pfam: PF00271; helicase\_C; 1.

DR SMART: SM00487; DEAD\_C; 1.  
 DR SMART: SM00490; HELIC\_C; 1.

DR PROSITE: PS00039; DEAD\_ATP\_HELICASE; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; UNKNOWN\_1.

KW ATP-binding; Helicase; RNA-binding.

SO SEQUENCE 578 AA; 62829 MW; 7B4A17CFD243D0E9 CRC64;

Query Match 63.8%; Score 37; DB 5; Length 578;  
 Best Local Similarity 70.0%; Pred. No. 1.7e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 KFGGGAAGG 11  
 DB 545 RFGGGAAGG 554

## RESULT 12

O90WR5 PRELIMINARY; PRT; 629 AA.

AC 090WR5; PRELIMINARY; PRT; 629 AA.  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
 DE KERATIN ALPHA.

OS Lampetra fluviatilis (River lamprey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;  
 OC Petromyzontiformes; Petromyzontidae; Lampetra.

NCBI\_TaxID=7748;

[1] SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;  
 MEDLINE-20196006; PubMed-10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA "unusual intermediate filament protein from Lampetra fluviatilis,  
 RT keratin-alpha."  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ289859; CAC42512.2; -  
 SQ SEQUENCE 629 AA; 63326 MW; A64975CC3566F9CB CRC64;

Query Match 63.8%; Score 37; DB 13; Length 629;  
 Best Local Similarity 54.5%; Pred. No. 1.8e+02;  
 Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKFGMGAKK 11  
 :||||:|  
 DB 39 SKFGVGILRG 49

## RESULT 13

O44356 PRELIMINARY; PRT; 131 AA.  
 O44356;

DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)  
 DE ABDUCTIN  
 OS Aequipecten irradians (Bay scallop).  
 OC Eukaryota; Metazoa; Mollusca; Bivalvia; Pteriomorphia; Pectinoida;  
 OC Pectinoidea; Pectinidae; Argopecten.  
 OX NCBI\_TaxID=31199;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98044338; Pubmed=9382816;  
 RA Cao Q.-P., Wang Y.-Q., Bayley H.;  
 RT "Sequence of abductin, the molluscan 'rubber' protein."  
 RL Curr. Biol. 7:0-0(1997).  
 DR EMBL; AF026847; AAB94679.1; -  
 SQ SEQUENCE 131 AA; 11034 MW; 3AF7BBB2A6BF3B84 CRC64;

Query Match 62.1%; Score 36; DB 5; Length 131;  
 Best Local Similarity 77.8%; Pred. No. 52;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKK 11  
 :|||||  
 DB 78 FCGMAKGG 86

## JULT 14

O94DD1 PRELIMINARY; PRT; 366 AA.  
 O94DD1;

DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE P0683F02.21 PROTEIN.  
 GN P0683F02.21.  
 OS Oryza sativa (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OX NCBI\_TaxID=4530;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. NIPPONBARE;  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC  
 clone:P0683F02."  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AP003289; BAB63718.1; -  
 SQ SEQUENCE 366 AA; 40220 MW; 4EBFADF20D9DB8 CRC64;

Query Match 62.1%; Score 36; DB 10; Length 366;

Best Local Similarity 66.7%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 3 FCGMGAKK 11  
 :||||:|  
 DB 176 FCGVGAESG 184

## RESULT 15

O58754 PRELIMINARY; PRT; 387 AA.  
 O58754;

DT 01-AUG-1998 (TREMBLrel. 07, Created)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE 387AA LONG HYDROTHERMAL AMIDOHYDROLASE.  
 GN PH1043.  
 OS Pyrococcus horikoshii.  
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
 OX NCBI\_TaxID=53953;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=OT3;  
 RX MEDLINE=98344137; Pubmed=9679194;  
 RA Kawarabayashi Y., Sawada M., Horikawa H., Hainaka Y., Hino Y.,  
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuku Y.,  
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,  
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
 RA Masuchi Y., Shizuya H., Kikuchi H.;  
 RT "Complete sequence and gene organization of the genome of a hyper-  
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."  
 RL DNA Res. 5:55-76(1998).  
 DR EMBL; AP000004; BAA30141.1; -  
 DR MEROPS; M40.002; -  
 DR InterPro; IPR002933; Peptidase\_M20.  
 DR Pfam; PF01546; Peptidase\_M20; 1.  
 KW Hydrolase; Complete proteome.  
 SQ SEQUENCE 387 AA; 43058 MW; 63DF19EEDCECE22 CRC64;

Query Match 62.1%; Score 36; DB 17; Length 387;  
 Best Local Similarity 70.0%; Pred. No. 1.6e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 AKFGMGAKK 10  
 :|||||  
 DB 136 AEEGGILGAKK 145

Search completed: October 28, 2002, 17:24:06  
 Job time : 28.4737 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:15 : Search time 22.3158 Seconds  
(without alignments)  
39.819 Million cell updates/sec

Title: US-09-833-079-1

Perfect score: 43  
Sequence: 1 PQGGGKVT 8

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A.GeneSeq.032802.\*  
1: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.\*  
2: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.\*  
3: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.\*  
4: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.\*  
5: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.\*  
6: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.\*  
7: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.\*  
8: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.\*  
9: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.\*  
10: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.\*  
11: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.\*  
12: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.\*  
13: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.\*  
14: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.\*  
15: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.\*  
16: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.\*  
17: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.\*  
18: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.\*  
19: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.\*  
20: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.\*  
21: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.\*  
22: /SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	24	11 AAR03021	N-terminal sequence of the to
2	43	100.0	161	10 AAP91911	Sequence of the Ga
3	43	100.0	6	AAP50045	E.coli HUB49 Gal-G
4	43	100.0	163	6 AAP60247	Pinin protein, Pap
5	43	100.0	163	22 AAB47081	Bacterial P pilus
6	43	100.0	163	22 AAB72835	Polypeptide sequen
7	35	81.4	312	22 AAU07897	Human gastric can
8	34	79.1	50	22 AAB63481	Human 5' EST relat
9	34	79.1	123	21 AAY65252	C glutamicum prote
10	34	79.1	143	22 AAG92121	Human histone H2A.
11	34	79.1	194	22 AAB73702	

12	34	79.1	486	22	ABG17822
13	34	79.1	896	22	AAU00820
14	34	79.1	2434	22	AAB42264
15	34	79.1	2594	22	AAU00984
16	32	74.4	66	22	ABG21990
17	32	74.4	109	21	AAU06272
18	32	74.4	109	21	AAU06272
19	32	74.4	172	21	AAU06271
20	32	74.4	172	21	AAU06271
21	32	74.4	172	21	AAU06271
22	32	74.4	173	21	AAU06270
23	32	74.4	173	21	AAU06270
24	32	74.4	173	21	AAU06270
25	32	74.4	185	21	AAU06270
26	32	74.4	306	16	AAU06270
27	32	74.4	405	22	AAU06270
28	32	74.4	513	16	AAU06270
29	32	74.4	513	20	AAU06270
30	32	74.4	522	19	AAU06270
31	32	74.4	548	20	AAU06270
32	32	74.4	596	20	AAU06270
33	32	74.4	1040	19	AAU06270
34	32	74.4	34	21	AAU06270
35	31	72.1	40	21	AAU06270
36	31	72.1	84	21	AAU06270
37	31	72.1	132	22	AAU06270
38	31	72.1	137	22	AAU06270
39	31	72.1	161	22	AAU06270
40	31	72.1	179	22	AAU06270
41	31	72.1	179	22	AAU06270
42	31	72.1	181	21	AAU06270
43	31	72.1	202	21	AAU06270
44	31	72.1	205	21	AAU06270
45	31	72.1	208	22	AAU06270

#### ALIGNMENTS

RESULT 1  
AAR03021  
AAR03021 standard: protein; 24 AA.  
AC AAR03021;  
DT 11-JUL-1990 (first entry)  
DE N-terminal sequence of Pglabo receptor specific protein.  
KW En; Dr; DrcMR; Pglabo; tissue receptor; blood typing.  
OS Escherichia coli.  
PN US4882425-A.  
PD 21-DEC-1989.  
PF 09-JUL-1987; 87US-0072197.  
PR 09-JUL-1987; 87US-0072197.  
PA (BAYU ) BAYLOR COLLEGE OF M.  
PI Hull RA, Hull SI, Nowicki B;  
PI WPI, 1990-036830/05.  
PT Receptor specific protein compns. Isolated from E.coli -  
PT useful for blood and tissue typing and for tissue staining.  
PS Claim 1; Page 18; 10pp; English.  
CC Protein has antigenic specificity for blood and tissue surface antigens.

CC May be used in blood and tissue typing for transfusion/transplantation,  
 CC paternity testing, disease diagnosis and tissue staining.  
 XX  
 SQ Sequence 24 AA;

Query Match 100.0%; Score 43; DB 11; Length 24;  
 Best Local Similarity 100.0%; Pred. No. 0.32;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 5 POGGKVT 12

RESULT 2  
 AAP91911  
 ID AAP91911 standard; peptide; 161 AA.  
 XX  
 AC AAP91911:

13-MAY-1990 (first entry)

Sequence of the total Fli-type subunit protein of Escherichia coli  
 DE fimbriae.

XX E.coli septicaemia; Fli-type fimbriae; colibacillosis;  
 KM CH4 wild type E. coli strain (02:K1:H-serotype).

XX Escherichia coli.

OS EP314224-A.

PN 03-MAY-1989.

PD 11-OCT-1988; 88EP-0202268.

PF 26-OCT-1987; 87NL-0002536.

PR (ALKU) AKZO NV.

PA van den Bosch J;

PI WPI; 1989-131778/18.

XX Vaccine for protecting poultry against E.coli septicaemia

PT -contg. Fli-type fimbriae or immunogenic sections, or

PT antibodies against these

Fig 1; page 11; 14pp; English.

XX An actual mol. wt. of 16.4kD can be calculated from the sequence.  
 CC Fli-type fimbriae was purified from CH4 (02:K1:H-serotype) E. coli  
 CC isolated from the affected hearts of chickens with colibacillosis. The  
 CC first 17 amino acids at the N-terminal are identical to the fli subunit.  
 CC A vaccine prepd. was prepared from Fli-type fimbriae which protected  
 CC poultry from E.coli septicaemia.

XX Sequence 161 AA;

Query Match 100.0%; Score 43; DB 10; Length 161;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 5 POGGKVT 12

RESULT 3  
 AAP50045  
 ID AAP50045 standard; protein; 163 AA.  
 XX  
 AC AAP50045;

XX 09-SEP-1991 (first entry)  
 DT Sequence of the Gal-gal-pilus protein.  
 DE Urinary tract infection; pilus vaccine; urinary pathogen.  
 XX  
 KM Escherichia coli H0849.  
 XX

Key Location/Qualifiers

FT 15..70  
 FT /label= region of antigenic specificity  
 FT /note= "can be isolated as a hydroxylamine II  
 FT fragment"

FT 79..110  
 FT /label= region of antigenic specificity  
 FT /note= "sequence corresponding to the tryptic IX  
 FT fragment"

FT 111..125  
 FT /label= region of antigenic specificity  
 FT /note= "sequence corresponding to the tryptic X  
 FT fragment"

FT 133..163  
 FT /label= region of antigenic specificity  
 FT /note= "isolated as a CNBr-HFBA III fragment"

PN EP161095-A.

PD 13-OCT-1985.

PF 29-APR-1985; 85EP-0303016.

PR 30-APR-1984; 84US-0605287.

PA (STRD) LELAND STANFORD JR UNIV.

PI O'Hanley P, Falkow S, Schoolnik G, Lark D;

DR WPI; 1985-284718/46.

XX Vaccines against urinary tract infections - contg. new E. coli

PT gal-gal pilus protein or fragments

XX Claim 7; Page 24; 29pp; English.

XX Methods for the prodn. of vaccines using AAP50045 and each of the

CC specified antigenic regions (see FT) are individually claimed.

CC AAP50045 and its fragments are highly effective and specific in

CC generating antibodies to urinary pathogens, and are obtainable in

CC practical amts. and in pure form.

XX Sequence 163 AA;

Query Match 100.0%; Score 43; DB 6; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 5 POGGKVT 12

RESULT 4  
 AAP60247  
 ID AAP60247 standard; protein; 163 AA.  
 XX  
 AC AAP60247;  
 XX 30-JUL-1991 (first entry)  
 DT E.coli H0849 Gal-gal pilus protein.  
 XX  
 DE H0849; pilin; vaccine; urinary tract infection;ss.  
 KM

XX OS Escherichia coli.  
XX PM EPI70496-A.  
XX PD 05-FEB-1986.  
XX PF 24-JUL-1985; 85EP-0305279.  
XX PR 30-JUL-1984; 84US-0635429.  
XX PA (STRD ) LELAND STANFORD JR UNIV.  
XX PI Schmidt MA, Shoolnik K, Ohanley P;  
XX DR WPI; 1986-037198/06.  
XX PT New peptide(s) having 8 or 11 amino acid residues - useful with  
XX PT derivs. as conjugates in vaccines for conferring protection  
XX PT against urinary tract infections  
XX PT  
XX Disclosure; Fig. 1; 24pp: English.  
XX  
CC Vaccines containing this protein confer protection against urinary  
CC infections, so that the need for antibiotic therapy is avoided. The  
CC protein is from a specific type of pilus associated with most.  
CC CC uropathogenic E. coli strains. The peptide fragments Pro(5)-Thr(12)  
CC and Ala(65)-Gly(75), each optionally having a Cys residue at one of  
CC their termini, are separately claimed. Any protein whose amino  
CC acid sequence includes one of these fragments is suitable for use  
CC in a vaccine, esp. the CNBr-II fragment corresponding to amino acids  
CC 53-163 of HU849 pilin.  
XX  
SQ Sequence 163 AA;

Query Match 100.0%; Score 43; DB 7; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
DB 5 POGGKVT 12

RESULT 5  
AAB47081  
ID AAB47081 standard; Protein; 163 AA.  
XX  
XX AAB47081;

08-MAY-2001 (first entry)

Pilin protein, Papa.

Donor; pilus protein; pilin; adhesin; vaccine; urinary epithelia;  
urinary tract infection; enterobacteriaceae.

Escherichia coli.

WO200104148-A2.

18-JAN-2001.

13-JUL-2000; 2000WO-US19066.

13-JUL-1999; 99US-0143582.

16-JUL-1999; 99US-0144359.

23-FEB-2000; 2000US-0184442.

(MEDI-) MEDIMUNE INC.

Hultgren SJ, Pinkner JS, Sauer F, Barnhart M, Waksman G, Knight S;

DR WPI; 2001-138315/14.  
XX  
XX Immunogenic complexes and polypeptides for vaccinating against urinary  
XX tract disease, comprises a pilus protein component and a bacterial  
XX chaperone -  
XX  
XX  
XX  
XX Disclosure; Fig 5; 92pp: English.

CC This sequence may be used as the pilus protein in the immunogenic  
CC complex of the invention. The complex comprises a pilus protein  
CC component and a donor complement portion as part of the same amino  
CC acid sequence or as non-covalently linked fragments of a complex  
CC such that the correct conformation of the pilin is maintained. The  
CC pilus protein component may be an adhesin or a pilin. pilus associated  
CC adhesins, such as FimH are relatively conserved proteins among  
CC different species and strains of bacteria, therefore vaccines  
CC incorporating the FimH antigen exhibit a broad spectrum of  
CC protection compared with current pilus-fiber based vaccines. The  
CC immunogenic complexes act by disrupting pilus-mediated attachment  
CC of E. coli to urinary epithelia and may prevent or retard the  
CC development of urinary tract infections. Vaccines containing the  
CC complexes are useful for preventing urinary tract disease in a human  
CC caused by the bacterium family enterobacteriaceae specifically  
CC Escherichia coli and may also be used in treating the disease.  
XX

SQ Sequence 163 AA;

Query Match 100.0%; Score 43; DB 22; Length 163;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
DB 5 POGGKVT 12

RESULT 6  
AAB72835  
ID AAB72835 standard; Protein; 163 AA.  
XX  
XX AAB72835;

09-MAY-2001 (first entry)

Bacterial P pilus subunit PapK.

Antibacterial compound; Gram-negative bacterium; pilus; chaperone;  
biofilm; disease treatment; bacterial infection.

Bacteria.

WO200110386-A2.

15-FEB-2001.

11-AUG-2000; 2000WO-US22087.

11-AUG-1999; 99US-0148280.

(UNITM ) UNIV WASHINGTON.

Hultgren SJ, Sauer FG, Waksman G, Fuetterer K;

WPI; 2001-226496/23.

An isolated compound for inhibiting pilus assembly -

Example 1; Fig 3b; 144pp: English.

CC The present invention provides antibacterial compounds which are able to  
CC interfere with Gram-negative bacteria pilus formation and assembly, and  
CC pilus interaction with chaperone proteins. These are useful in the  
CC treatment of bacterial infection, and in the prevention of biofilm

CC formation. They are particularly useful against *Escherichia coli*,  
 CC *Haemophilus influenzae*, *Salmonella enteritidis*, *S. typhimurium*,  
 CC *Bordetella pertussis*, *Yersinia enterocolitica*, *Helicobacter pylori* and  
 CC *Klebsiella pneumoniae*.

XX  
 SQ Sequence 163 AA;

Query Match 100.0%; Score 43; DB 22; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 2.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 5 POGGKVT 12

RESULT 7  
 AAU07897  
 ID AAU07897 standard; Protein: 312 AA.

AAU07897;

DT 18-DEC-2001 (first entry)

DE Polypeptide sequence for human hspg39b.

XX Human; reproductive-specific protein; male infertility; gene therapy;  
 KW spermatogenesis; sperm count disorder; anti infertility; reproduction.

XX Homo sapiens.

PN WO20016752-A2.

PD 13-SEP-2001.

PF 07-MAR-2001; 2001WO-US07371.

PR 07-MAR-2000; 2000US-0187518.

PR 12-JAN-2001; 2001US-0261557.

PA (WHED ) WHITEHEAD INST BIOMEDICAL RES.

PI Wang PJ, Page DC;

DR WPI; 2001-570774/64.

DR N-PSDB; AAS13663.

XX Novel reproduction-specific protein, useful for treating disorders of  
 reduced sperm count, enhancing/increasing sperm count and/or sperm  
 activity -

PS Claim 22; Fig 85; 151pp; English.

XX The present invention relates to the isolation of novel mammalian and  
 CC human reproductive-specific proteins (AAU07859-AAU07899), and the  
 CC nucleic acids encoding them. The nucleic acids encoding  
 CC reproductive-specific proteins are useful for diagnosing infertility  
 CC which is a result of reduced sperm count, reduced sperm motility,  
 CC malformed sperm or combinations of these. The sequences of the invention  
 CC are useful as markers for spermatogonial cells, for identifying genes or  
 CC proteins characteristic of male infertility, diagnosing or aiding in  
 CC the diagnosis of infertility in men, and for contraception in which  
 CC sperm production or sperm count is reduced or defective sperm is  
 CC produced. Antibodies to reproductive-specific proteins are useful for  
 CC determining the presence of these proteins in a sample obtained from a  
 CC man being assessed for infertility, for identifying the expression of  
 CC genes in particular cell type or particular developmental stage, for  
 CC studies of spermatogenesis, and for immunofluorescence of germ cells or  
 CC in Western blots for assessing the presence of the protein the antibody  
 CC binds. The sequences of the invention are also useful for treating  
 CC disorders of reduced sperm count, and for increasing sperm count and/or  
 CC sperm activity. The nucleic acids of the invention are useful in gene  
 CC therapy. AAU07883-AAU07899 represent the human reproduction-specific

CC proteins of the present invention.

XX  
 SQ Sequence 312 AA;

Query Match 81.4%; Score 35; DB 22; Length 312;  
 Best Local Similarity 75.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 158 POGGQCAT 165

RESULT 8  
 AAB63481  
 ID AAB63481 standard; Protein: 50 AA.

AC AAB63481;

DT 26-MAR-2001 (first entry)

DE Human gastric cancer associated antigen protein sequence SEQ ID NO:843.

XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;  
 KW cancer associated antigen; cytostatic; cancer vaccine.

XX Homo sapiens.

PN WO200073801-A2.

PD 07-DEC-2000.

PF 26-MAY-2000; 2000WO-US14749.

PR 28-MAY-1999; 99US-0136526.

PR 10-SEP-1999; 99US-0153454.

PA (LUDW-) LUDWIG INST CANCER RES.

PI Obata Y;

DR WPI; 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated  
 PT antigen precursors, useful for diagnosing and treating a condition  
 PT characterized by expression of an abnormal amount of a protein, e.g.  
 PT cancer -

XX Example 1; Page 583; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014  
 CC represent nucleotide sequences encoding human breast, gastric and  
 CC prostate cancer associated antigen precursors (CAAP) respectively.  
 CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970  
 CC represent human breast, gastric and prostate CAAP protein sequence  
 CC respectively. CAAPs have cytostatic activity and can be used in the  
 CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic  
 CC acids or anti-CAAP antibodies are useful for diagnosing and treating a  
 CC condition characterised by expression of an abnormal amount of a protein,  
 CC e.g. cancer.

XX Sequence 50 AA;

Query Match 79.1%; Score 34; DB 22; Length 50;  
 Best Local Similarity 100.0%; Pred. No. 31;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGK 6  
 |||||  
 Db 1 POGGK 6

RESULT 9

```

AA65252
ID      AA65252 standard; Protein; 123 AA.
XX
XX      AA65252;
AC
DT      01-FEB-2000 (first entry)
XX
DE      Human 5' EST related polypeptide SEQ ID NO:1413.
XX
KW      Human; 5' EST; expressed sequence tag; secreted protein; diagnosis;
KW      gene therapy; chromosome mapping; upstream regulatory sequence;
KW      forensic; location; development; protein synthesis; stability;
XX      regulation; identification.
OS      Homo sapiens.
PN      MO9953051-A2.
PD      21-OCT-1999.
XX
XX      09-APR-1999; 99WO-IB00712.
PR      09-APR-1998; 98US-0057719.
PR      28-APR-1998; 98US-0069047.
PA      (GEST ) GENSET.
PI      Dunas Milne Edwards J, Duclert A, Giordano J;
DR      WPI: 2000-038446/03.
DR      N-PSDB: AA242866.
XX
PT      Novel secreted protein 5' expressed sequence tag sequences used in
PT      diagnostic, forensic, gene therapy, and chromosome mapping procedures
XX
PS      Claim 3; Page 777; 837pp; English.
XX
CC      AA42265 to AA243075 represent novel 5' expressed sequence tag (EST)
CC      sequences, corresponding to human secreted proteins. AA64651 to
CC      AA6538 represent the EST-related proteins corresponding to AA42265 to
CC      AA43052. The 5' ESTs can be used for producing secreted human gene
CC      products. They can be used to identify and isolate 5' untranslated
CC      regions (UTRs) and upstream regulatory regions which control the
CC      location, development stage, rate, and quantity of protein synthesis, as
CC      well as stability of mRNA. The ESTs are also useful as probes for
CC      chromosome mapping, and to obtain full length cDNA clones. The ESTs can
CC      also be used in forensic procedures to identify individuals, or in
CC      diagnostic procedures to identify individuals having genetic diseases
CC      resulting from abnormal gene expression. The products may also be used in
CC      gene therapy protocols. The nucleic acids encoding signal peptides can be
CC      used for directing extracellular secretion of a polypeptide or the
CC      insertion of a polypeptide into a membrane, or importing a polypeptide
CC      into a cell. The proteins encoded by the EST sequences may be useful in
CC      treating a variety of human conditions. Secreted proteins have
CC      therapeutic value, and the identification of new secreted proteins is
CC      valuable. AA42249 to AA42264 and AA64644 to AA64650 represent
CC      sequences used in the exemplification of the present invention.
XX
SO      Sequence 123 AA;

Query Match          79.1%; Score 34; DB 21; Length 123;
Best Local Similarity 62.5%; Pred. No. 76;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY      1 POGSGKVT 8
        ||| |::|
DB      26 POGSGRIT 33

RESULT 10
ID      AA692121
XX      AA692121 standard; Protein; 143 AA.
XX
```

AC		AA692121;	
XX			
DT		26-SEP-2001	(first entry)
XX			
DE	C	glutamicum protein fragment SEQ ID NO: 5875.	
XX			
KW	Corynebacterium bacterium; amino acid synthesis; vitamin; saccharide;		
XX	organic acid synthesis.		
OS			
PN	Corynebacterium glutamicum.		
XX			
PD	EPI108790-A2.		
XX			
PF	20-JUN-2001.		
XX			
PR	18-DEC-2000; 2000EP-0127688.		
XX			
PR	16-DEC-1999; 99JP-0377484.		
XX	07-APR-2000; 2000JP-0159162.		
PR	03-AUG-2000; 2000JP-0280988.		
XX			
PA	(KYOW ) KYOMA HAKKO KOGYO KK.		
XX			
PI	Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;		
PI	Tateishi N, Senoh A, Ikeda M, Ozaki A;		
XX			
DR	WPI: 2001-376931/40.		
DR	N-PSDB: AAH67340.		
XX			
PT	Novel polynucleotides derived from Corynebacterium bacteria, for identifying		
PT	mutation point of a gene, measuring expression of a gene, analysing		
PT	expression profile or pattern of a gene and identifying homologous gene		
XX			
PS	Claim 17; SEQ ID NO: 5875; 246pp + Sequence Listing; English.		
XX			
CC	The present invention provides a number of nucleotide and protein		
CC	sequences from the Corynebacterium bacterium Corynebacterium glutamicum. These		
CC	are useful for identifying the mutation point of a gene derived from a		
CC	mutant of coryneform bacterium, measuring expression amount and		
CC	analysing the expression profile or expression pattern of a gene derived		
CC	from coryneform bacterium, and identifying a homologue of a gene derived		
CC	from coryneform bacterium. Corynebacterium bacteria are useful for producing		
CC	amino acids, nucleic acids, vitamins, saccharides and organic acids,		
CC	particularly L-lysine. The present sequence is a protein described		
CC	in the exemplification of the invention.		
CC	Note: The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from the		
XX	European Patent Office.		
SQ	Sequence 143 AA:		
Query Match	79.1%; Score 34; DB 22; Length 143;		
Best Local Similarity	62.5%; Pred.No. 88;		
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;			
OY	1 PQGGGKVT 8		
I I I I I I I			
Db	81 PDGGGRIT 88		
RESULT 11			
ID	AA673702		
AB	AA673702 standard; Protein; 194 AA.		
AC	AA673702;		
AD			
AE	11-SEP-2001 (first entry)		
AF			
AG	Human histone H2A.21.		
AH			
AI			
AJ	Histone H2A.21; human; histone H2A.1 homologue; recombinant production;		
AK	malignant tumour; cancer; blood disease; HIV infection;		
AL			

KW human immunodeficiency virus; immune disorder; inflammatory condition;  
KM cytostatic; anti-HIV; antiinflammatory; immunomodulator.  
XX  
OS Homo sapiens.  
XX  
PN WO200138522-A1.  
PD  
XX 31-MAY-2001.  
XX  
PF 20-NOV-2000; 2000MO-CN00440.  
XX  
PR 23-NOV-1999; 99CN-0124083.  
XX  
PA (BIO-R-) BIOROAD GENE DEV LTD SHANGHAI.  
XX  
PI Mao Y, Xie Y;  
XX  
DR WPI; 2001-355930/37.  
XX  
N-PSDB; AAH24271.  
XX  
PT New human histone H2A.21 for diagnosing and treating malignant tumor,  
PT hemopathy, human immunodeficiency virus (HIV) infection, immunological  
PT diseases and inflammation -  
XX  
PS Claim 1; Page 20; 31pp; Chinese.  
XX  
CC This sequence represents human histone H2A.21. The protein has a  
CC molecular weight of 21 kd, and has 98% identity and 100% homology  
CC with human histone H2A.1 (GenBank accession number M60752) over an 84  
CC amino acid stretch. The invention relates to human histone H2A.21  
CC (AA03702), nucleic acids encoding it (AAH24271), and a method for the  
CC recombinant production of human histone H2A.21. The present invention  
CC additionally discloses an antibody which specifically binds to human  
CC histone H2A.21. Human histone H2A.21, and nucleotides which encode it  
CC may be used for treating a variety of diseases, such as malignant  
CC tumours, blood diseases, HIV (human immunodeficiency virus) infection,  
CC immune disorders and inflammatory conditions. The protein may also be  
CC used to screen for modulators of its activity or for peptide  
CC fingerprinting identification. The polynucleotide can be used as a  
CC primer for nucleic acid amplification reaction or as a probe for  
CC hybridisation reactions, or in producing gene chips or microarrays.  
XX  
SQ Sequence 194 AA;  
XX  
Query Match 79.1%; Score 34; DB 22; Length 194;  
Best Local Similarity 85.7%; Pred. No. 1.2e+02;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
XX  
1 POGGKGV 7  
|||||  
Db 125 POGGGOV 131  
XX  
RESULT 12  
ABG17822  
ID ABG17822 standard; Protein; 486 AA.  
XX  
AC ABG17822;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #17813.  
XX  
KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KM food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.

XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Dermanac RT, Liu C, Tang YF;  
XX  
DR WPI; 2001-639362/73.  
XX  
DR N-PSDB; AAS82009.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID No 48181; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pt\_sequences.  
XX  
SQ Sequence 486 AA;  
XX  
Query Match 79.1%; Score 34; DB 22; Length 486;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 POGGK 6  
|||||  
Db 34 POGGK 39  
XX  
RESULT 13  
AAM00820  
ID AAM00820 standard; Protein; 896 AA.  
XX  
AC AAM00820;  
XX  
DT 01-OCT-2001 (first entry)  
XX  
DE Human bone marrow protein, SEQ ID NO: 183.  
XX  
KM Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;  
KM antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
KM immunosuppressive; gene therapy; cytokine cell proliferation;  
KM cell differentiation modulator; immune disorder; infection; cancer;  
KM human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.  
XX  
OS Homo sapiens.  
XX  
PN WO200153453-A2.  
XX  
PD 26-JUL-2001.  
XX

PF 23-DEC-2000; 2000WO-US34960.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 30-NOV-2000; 2000US-0250583.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;  
 PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Drmanac RT;  
 DR WPI; 2001-488707/53.  
 DR N-PSDB; AAH89939.  
 XX  
 PS Novel bone-marrow-expressed polynucleotides and polypeptides, useful  
 for treating e.g. cancer and immune deficiency disorders -  
 CC  
 PS Claim 10; Page 338-339; 648pp; English.  
 CC  
 CC The present sequence is one of 251 novel human polypeptides encoded  
 by a bone marrow-expressed polynucleotide. The polynucleotide and the  
 polypeptide encoded by it are useful in the treatment of various  
 CC immune deficiencies and disorders. The deficiencies and disorders may  
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal  
 CC infection, or may result from an autoimmune disorder, a coagulation  
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,  
 CC suppression of an inflammatory response or treatment of a nervous  
 CC system disorder such as Alzheimer's disease. Detection of the presence  
 CC or increased expression of the polynucleotide or the protein it  
 CC encodes is useful for the diagnosis and/or prognosis of one  
 CC or more types of cancer. The polynucleotide and polypeptide can be  
 CC used as nutritional sources or supplements and in the screening of  
 CC chemical compounds as potential drugs.  
 CC  
 XX  
 SQ Sequence 896 AA;  
 Query Match 79.1%; Score 34; DB 22; Length 896;  
 Best Local Similarity 75.0%; Pred. No. 5.7e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 POGGCKVT 8  
 |||:|  
 277 POGECTVT 284  
 RESULT 14  
 AAB42264  
 ID AAB42264 standard; Protein: 2434 AA.  
 XX  
 AC AAB42264;  
 XX  
 DT 08-FEB-2001 (first entry)  
 XX  
 DE Human ORFX ORF2028 polypeptide sequence SEQ ID NO:4056.  
 XX  
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;  
 KW vulnerrary; antiparotiatric; antiparkinsonian; neurotropic; neuroprotective;  
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;  
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;  
 KW hypotensive; dermatological; immunosuppressive; antinflammatory;  
 KW antiviral; antibacterial; antifungal; antineumatic; antihydroid;  
 KW antinaemic; gene therapy; cancer; proliferative disorder; hypertension;  
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;  
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;  
 KW cholesterol ester storage; systemic lupus erythematosus; infection;  
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;  
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;

KW bone damage; cartilage damage; antinflammatory disease; coagulation;  
 KW thrombosis; contraceptive.  
 KW  
 OS Homo sapiens.  
 XX  
 PN WO200058473-A2.  
 XX  
 XX 05-OCT-2000.  
 XX  
 PD 31-MAR-2000; 2000WO-US08621.  
 PF  
 XX 31-MAR-1999; 99US-0127607.  
 PR 02-APR-1999; 99US-0127636.  
 PR 05-APR-1999; 99US-0127728.  
 PR 30-MAR-2000; 2000US-0540763.  
 XX  
 XX (CURA-) CURAGEN CORP.  
 XX  
 XX Shinkets RA, Leach M;  
 PI WPI; 2000-602362/57.  
 DR N-PSDB; AAC76473.  
 DR  
 XX  
 PS Claim 11; Page 3239-3244; 5507pp; English.  
 PS  
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,  
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX  
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerrary;  
 CC antiparotiatric; antiparkinsonian; neurotropic; neuroprotective;  
 CC anticonvulsant; osteopathic; antiarthritic; immunosuppressant;  
 CC osteopathic; anticonvulsant; antiparotiatric; immunosuppressant;  
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;  
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;  
 CC antinflammatory; antibacterial; antiviral; antifungal; antineumatic;  
 CC antihydroid; and antinaemic. The sequences can be used for determining  
 CC the presence of or predisposition to, or preventing or treating  
 CC pathological conditions associated with an ORFX-associated disorder. The  
 CC nucleic acids can be used to express ORFX proteins in gene therapy  
 CC vectors. The proteins and nucleic acids may be used to treat cancers,  
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,  
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,  
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus  
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,  
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,  
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,  
 CC nocturnal haemoglobinuria, antinflammatory disease; to enhance  
 CC coagulation; to inhibit thrombosis; and as a contraceptive.  
 CC  
 XX  
 SQ Sequence 2434 AA;  
 Query Match 79.1%; Score 34; DB 21; Length 2434;  
 Best Local Similarity 75.0%; Pred. No. 1.6e+03;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 OY 1 POGGCKVT 8  
 |||:|  
 DB 1119 POGECTVT 1126  
 RESULT 15  
 AAM00984  
 ID AAM00984 standard; Protein: 2594 AA.  
 XX  
 AC AAM00984;  
 XX  
 DT 01-OCT-2001 (first entry)  
 XX  
 DE Human bone marrow protein, SEQ ID NO: 485.  
 XX  
 KW Human; bone marrow; antinflammatory; cytostatic; neuroprotective;

KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;  
 KW immunosuppressive; gene therapy; cytokine cell proliferation;  
 KW cell differentiation modulator; immune disorder; infection; cancer;  
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.

OS Homo sapiens.

PN WO200153453-A2.

PD 26-JUL-2001.

PF 23-DEC-2000; 2000WO-US34960.

PR 21-JAN-2000; 2000US-0488725.

PR 25-APR-2000; 2000US-0552317.

PR 09-JUL-2000; 2000US-0598042.

PR 19-JUL-2000; 2000US-0620312.

PR 03-AUG-2000; 2000US-0653450.

PR 14-SEP-2000; 2000US-0662191.

PR 19-OCT-2000; 2000US-0693036.

PR 30-NOV-2000; 2000US-0250583.

PI (HYSE-) HYSEQ INC.

PI Ford JE, Boyle BJ, Tang YF, Liu C, Asundi V, Chen R, Ma Y;  
 PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Drmanac RT;

PI WPI: 2001-48707/53.

PI N-PSDB; AAH90103.

PI Novel bone-marrow-expressed polynucleotides and polypeptides, useful  
 PI for treating e.g. cancer and immune deficiency disorders -

PS Claim 10: Page 622-627; 648pp; English.

CC The present sequence is one of 251 novel human polypeptides encoded  
 CC by a bone marrow-expressed polynucleotide. The polynucleotide and the  
 CC polypeptide encoded by it are useful in the treatment of various  
 CC immune deficiencies and disorders. The deficiencies and disorders may  
 CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal  
 CC infection, or may result from an autoimmune disorder, a coagulation  
 CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,  
 CC suppression of an inflammatory response or treatment of a nervous  
 CC system disorder such as Alzheimer's disease. Detection of the presence  
 CC or increased expression of the polynucleotide or the protein it  
 CC encodes is useful for the diagnosis and/or prognosis of one  
 CC or more types of cancer. The polynucleotide and polypeptide can be  
 CC used as nutritional sources or supplements and in the screening of  
 CC chemical compounds as potential drugs.

SQ Sequence 2594 AA:

Query Match 79.1%; Score 34; DB 22; Length 2594;

Best Local Similarity 75.0%; Pred. No. 1.7e+03;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||:| |  
 Db 1279 POGGKVT 1286

Search completed: October 28, 2002, 17:22:38  
 Job time : 24.3158 secs



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 ; Search time 8.8421 seconds  
(without alignments)  
22.099 Million cell updates/sec

Title: US-09-833-079-1

Perfect score: 43  
Sequence: 1 POGGKVT 8

Scoring table:

BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

al number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCRNUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	74.4	513	4	US-09-100-193-3	Sequence 3, Appli
2	74.4	522	4	US-08-961-083-120	Sequence 120, App
3	74.4	1040	4	US-08-961-083-118	Sequence 118, App
4	72.1	237	4	US-08-961-083-136	Sequence 136, App
5	72.1	902	1	US-08-396-479B-6	Sequence 6, Appli
6	72.1	902	1	US-08-818-823-6	Sequence 6, Appli
7	69.8	311	2	US-08-484-905-70	Sequence 70, Appli
8	69.8	311	2	US-08-481-985B-70	Sequence 70, Appli
9	69.8	311	4	US-08-370-476-70	Sequence 70, Appli
10	69.8	328	2	US-08-828-242-1	Sequence 1, Appli
11	69.8	328	2	US-08-910-927B-1	Sequence 1, Appli
12	69.8	328	4	US-09-206-499-1	Sequence 1, Appli
13	69.8	328	4	US-09-270-270-1	Sequence 1, Appli
14	69.8	454	3	US-08-348-518C-4	Sequence 4, Appli
15	69.8	454	3	US-08-476-509B-4	Sequence 4, Appli
16	69.8	680	3	US-08-761-136-1	Sequence 1, Appli
17	69.8	796	2	US-08-817-900-2	Sequence 2, Appli
18	69.8	796	4	US-09-236-645-2	Sequence 2, Appli
19	67.4	85	3	US-08-772-440-32	Sequence 32, Appli
20	67.4	131	3	US-08-772-440-23	Sequence 23, Appli
21	67.4	131	3	US-08-772-440-27	Sequence 27, Appli
22	67.4	134	3	US-08-772-440-16	Sequence 16, Appli
23	67.4	145	3	US-08-772-440-14	Sequence 14, Appli
24	67.4	167	3	US-08-772-440-21	Sequence 21, Appli
25	67.4	175	3	US-08-772-440-15	Sequence 15, Appli
26	67.4	184	2	US-08-737-825-10	Sequence 10, Appli
27	67.4	209	3	US-08-772-440-4	Sequence 4, Appli

28	29	67.4	226	4	US-09-268-364-2	Sequence 2, Appli
29	29	67.4	256	1	US-08-594-808B-7	Sequence 7, Appli
30	29	67.4	631	4	US-08-971-188-8	Sequence 8, Appli
31	29	67.4	691	5	PCT-US91-08442-2	Sequence 2, Appli
32	29	67.4	778	4	US-09-460-145-2	Sequence 2, Appli
33	29	67.4	797	4	US-09-460-145-4	Sequence 4, Appli
34	29	67.4	913	3	US-08-827-208-3	Sequence 3, Appli
35	29	67.4	913	4	US-09-500-358-3	Sequence 3, Appli
36	29	67.4	913	4	US-09-498-809-3	Sequence 3, Appli
37	28	65.1	13	5	PCT-US94-10257A-51	Sequence 51, Appli
38	28	65.1	13	5	PCT-US94-10257A-52	Sequence 52, Appli
39	28	65.1	20	5	PCT-US94-10257A-80	Sequence 80, Appli
40	28	65.1	27	2	US-08-488-161-30	Sequence 30, Appli
41	28	65.1	27	3	US-09-273-685-30	Sequence 30, Appli
42	28	65.1	27	5	PCT-US95-11934-30	Sequence 30, Appli
43	28	65.1	73	4	US-09-063-950-9	Sequence 9, Appli
44	28	65.1	80	4	US-09-091-219-6	Sequence 6, Appli
45	28	65.1	122	3	US-08-480-173A-56	Sequence 56, Appli

## ALIGNMENTS

RESULT 1  
US-09-100-193-3  
Sequence 3, Application US/09100193  
Patent No. 6153729  
GENERAL INFORMATION:  
APPLICANT: Gary S. Stein et al.  
TITLE OF INVENTION: NUCLEAR MATRIX TARGETING PEPTIDES AND USES THEREFORE  
NUMBER OF SEQUENCES: 14  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD  
STREET: 28 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/100,193  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/050,104  
FILING DATE: 20-JUNE-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: UMW-024  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)742-4214  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 513 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-09-100-193-3  
Query Match 74.4%: Score 32: DB 4: Length 513;  
Best Local Similarity 62.5%: Pred. No. 1.8e+02;  
Matches 5: Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
QY 1 POGGKVT 8  
||||:|  
Db 271 POGGKVT 278

```
RESULT 2
US-08-961-083-120
; Sequence 120, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8504
; INFORMATION FOR SEQ ID NO: 120:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 522 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-120

Query Match          74.4% Score 32; DB 4; Length 522;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGQGKVT 8
Db 131 QGQGKVS 137

RESULT 3
US-08-961-083-118
; Sequence 118, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
```

```
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 118:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1040 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-118

Query Match          74.4% Score 32; DB 4; Length 1040;
Best Local Similarity 85.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 QGQGKVT 8
Db 131 QGQGKVS 137

RESULT 4
US-08-961-083-136
; Sequence 136, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 136:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

MOLECULE TYPE: protein  
US-08-961-083-136

Query Match 72.1%; Score 31; DB 4; Length 237;  
Best Local Similarity 83.3%; Pred. No. 1.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 PQGGKV 6  
111111  
DB 117 PQGGKV 122

## RESULT 5

US-08-396-479B-6  
Sequence 6, Application US/08396479B

Patent No. 5612455

GENERAL INFORMATION:

APPLICANT: HOEY, Timothy

TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/396,479B

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59450-1/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771

TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 902 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-396-479B-6

Query Match 72.1%; Score 31; DB 1; Length 902;  
Best Local Similarity 85.7%; Pred. No. 4.9e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQGGKV 7  
111111  
DB 559 PQGGKV 565

## RESULT 6

US-08-818-823-6

Sequence 6, Application US/08818823

Patent No. 5708158

GENERAL INFORMATION:

APPLICANT: HOEY, Timothy

TITLE OF INVENTION: NUCLEAR FACTORS AND BINDING ASSAY

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California  
COUNTRY: USA

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,823

FILING DATE: 14-MAR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/396,479

FILING DATE: 02-MAR-1995

ATTORNEY/AGENT INFORMATION:

NAME: Osman, Richard A

REGISTRATION NUMBER: 36,627

REFERENCE/DOCKET NUMBER: A-59450-1/RAO

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 494-8700

TELEFAX: (415) 494-8771

TELEX: 210 277299

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 902 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-818-823-6

Query Match 72.1%; Score 31; DB 1; Length 902;  
Best Local Similarity 85.7%; Pred. No. 4.9e+02;

Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 PQGGKV 7  
111111  
DB 559 PQGGKV 565

## RESULT 7

US-08-484-905-70

Sequence 70, Application US/08484905

Patent No. 5976551

GENERAL INFORMATION:

APPLICANT: Mottez, Estelle

APPLICANT: Abastado, Jean-Pierre

APPLICANT: Kourilsky, Philippe

TITLE OF INVENTION: An Altered Major Histocompatibility

TITLE OF INVENTION: Complex(MHC) Determinant and Methods for Using the

NUMBER OF SEQUENCES: 127

CORRESPONDENCE ADDRESS:

ADDRESSEE: Finnegan, Henderson, Farbow, Garrett &

ADDRESSEE: Dunner

STREET: 1300 I Street, N.W., Suite 700

CITY: Washington

STATE: D.C.

ZIP: 20005-3315

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/484,905

FILING DATE: 07-JUNE-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/801,818

FILING DATE: 05-DEC-1991

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Potter, Jane E. R.  
REGISTRATION NUMBER: 33,332  
REFERENCE/DOCKET NUMBER: 03495.0106-03000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4400  
TELEFAX: 202-408-4000  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-484-905-70

Query Match  
Best Local Similarity 69.8%; Score 30; DB 2; Length 311;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGOKVT 8  
I: :|||||  
Db 193 PRSEKVT 200

RESULT 8  
US-08-481-985B-70  
Sequence 70, Application US/08481985B  
Patent No. 6011146  
GENERAL INFORMATION:  
APPLICANT: Motiez, Estelle  
APPLICANT: Abastado, Jean-Pierre  
APPLICANT: Kourilsky, Philippe  
TITLE OF INVENTION: Altered Major Histocompatibility Complex  
NUMBER OF SEQUENCES: 148  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,985B  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/801,818  
FILING DATE: 05-DEC-1991  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 03495.0106-04000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4400  
TELEFAX: 202-408-4000  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid

TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-481-985B-70

Query Match  
Best Local Similarity 69.8%; Score 30; DB 3; Length 311;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGOKVT 8  
I: :|||||  
Db 193 PRSEKVT 200

RESULT 9  
US-08-370-476-70  
Sequence 70, Application US/08370476  
Patent No. 6153408  
GENERAL INFORMATION:  
APPLICANT: Motiez, Estelle  
APPLICANT: Abastado, Jean-Pierre  
APPLICANT: Kourilsky, Philippe  
APPLICANT: Lone, Yu-Chun  
APPLICANT: Ojcius, David  
APPLICANT: Castrouge, Amanda  
TITLE OF INVENTION: Altered Major Histocompatibility Complex  
NUMBER OF SEQUENCES: 127  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &  
STREET: 1300 I Street, N.W., Suite 700  
CITY: Washington  
STATE: D.C.  
ZIP: 20005-3315  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/370,476  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/117,575  
FILING DATE: 07-SEP-1993  
APPLICATION NUMBER: US 08/072,787  
FILING DATE: 06-JUN-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/801,818  
FILING DATE: 05-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/792,473  
FILING DATE: 15-NOV-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Meyers, Kenneth J.  
REGISTRATION NUMBER: 25,146  
REFERENCE/DOCKET NUMBER: 05243.0001-01000  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-408-4000  
TELEFAX: 202-408-4400  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-370-476-70

Query Match  
Best Local Similarity 69.8%; Score 30; DB 4; Length 311;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
Db 193 PRSEKVT 200

## RESULT 10

US-08-828-242-1  
; Sequence 1, Application US/08828242  
; Patent No. 5871970  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL CALCIUM-BINDING  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/828, 242  
; FILING DATE: Filed Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-845-0555  
; TELEFAX: 415-845-4166  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 328 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: CONUTUT01  
; CLONE: 2509570  
; US-08-828-242-1

Query Match 69.8%; Score 30; DB 2; Length 328;  
Best Local Similarity 71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGGKVT 7  
Db 28 PHGGGRV 34

RESULT 11  
US-08-910-927B-1  
; Sequence 1, Application US/08910927B  
; Patent No. 5976801  
; GENERAL INFORMATION:  
; APPLICANT: Bandman, Olga  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi  
; TITLE OF INVENTION: HUMAN RETICULOCALBIN ISOFORMS  
; NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/910, 927B  
FILING DATE: Herewith  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:

ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0358 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 328 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: RATRNOT02  
CLONE: 922578  
US-08-910-927B-1

Query Match 69.8%; Score 30; DB 2; Length 328;  
Best Local Similarity 71.4%; Pred. No. 2.7e+02;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGGKVT 7  
Db 28 PHGGGRV 34

## RESULT 12

US-09-206-499-1  
; Sequence 1, Application US/09206499  
; Patent No. 6194385  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Goli, Surya K.  
; TITLE OF INVENTION: NOVEL CALCIUM-BINDING  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/206, 499  
; FILING DATE:  
; CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/828,242  
FILING DATE: 03/31/1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0261 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-845-0555  
TELEFAX: 415-845-4166  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 328 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: CONUTUT01  
CLONE: 2509570  
J-206-499-1

Query Match  
Best Local Similarity 69.8%; Score 30; DB 4; Length 328;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGQGRV 7  
| | | | |  
Db 28 POGQGRV 34

RESULT 13  
US-09-270-270-1  
Sequence 1, Application US/09270270  
Patent No. 6235477  
GENERAL INFORMATION:  
APPLICANT: Bandman, Olga  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Lal, Preeti  
APPLICANT: Corley, Neil C.  
APPLICANT: Shah, Purvi  
TITLE OF INVENTION: HUMAN RETICULOCALBIN ISOFORMS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Drive  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/270,270  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/910,927  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.  
REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0358 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-855-0555  
TELEFAX: 650-845-4166  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 328 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
LIBRARY: RATRNOT02  
CLONE: 922578  
US-09-270-270-1

Query Match  
Best Local Similarity 69.8%; Score 30; DB 4; Length 328;  
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGQGRV 7  
| | | | |  
Db 28 POGQGRV 34

RESULT 14  
US-08-348-518C-4  
Sequence 4, Application US/08348518C  
Patent No. 6022740  
GENERAL INFORMATION:  
APPLICANT: SUDOL, MARIUS  
APPLICANT: PIER, BORK  
APPLICANT: HENRY, CHEN  
TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 26  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Klauber & Jackson  
STREET: 411 Hackensack Avenue  
CITY: Hackensack  
STATE: New Jersey  
COUNTRY: USA  
ZIP: 07601  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/348,518C  
FILING DATE: 01-DEC-1994  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Jackson Esq., David A.  
REGISTRATION NUMBER: 26,742  
REFERENCE/DOCKET NUMBER: 600-1-101  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 201 487-5800  
TELEFAX: 201 343-1684  
TELEX: 133521  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 454 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-518C-4

Query Match  
Best Local Similarity 69.8%; Score 30; DB 3; Length 454;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGQGRV 6  
| | | | |  
Db 13 POGQGRV 18

RESULT 15  
US-08-476-509B-4  
Sequence 4, Application US/08476509B

; Patent No. 6034212  
; GENERAL INFORMATION:  
; APPLICANT: SUDOL, MARTINUS  
; APPLICANT: PEER, BORK  
; APPLICANT: HENRY, CHEN  
; TITLE OF INVENTION: A SH2 DOMAIN ASSOCIATED PROTEIN, A  
; TITLE OF INVENTION: SIGNALING DOMAIN THEREIN, NUCLEIC ACIDS ENCODING THE  
; TITLE OF INVENTION: PROTEIN AND THE DOMAIN, AND DIAGNOSTIC AND THERAPEUTIC USES  
; TITLE OF INVENTION: THEREOF  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klauber & Jackson  
; STREET: 411 Hackensack Avenue  
; CITY: Hackensack  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,509B  
; FILING DATE: 01-DEC-1994  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jackson Esq., David A.  
; REGISTRATION NUMBER: 26,742  
; REFERENCE/DOCKET NUMBER: 600-1-101 CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 487-5800  
; TELEFAX: 201 343-1684  
; TELEX: 133521  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 454 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-476-509B-4

Query Match 69.8%; Score 30; DB 3; Length 454;  
Best local similarity 83.3%; Pred. NO. 3.8e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 POGGCK 6  
|||||:  
13 POGGCK 18

Search completed: October 28, 2002, 17:24:30  
Job time : 9.8421 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 ; Search time 11.7895 Seconds  
(without alignments)  
65.203 Million cell updates/sec

Title: US-09-833-079-1

Perfect score: 43

Sequence: 1 PQGGGKVT 8

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

al number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43	100.0	161	2 S25206	prsa protein - Esc
2	43	100.0	185	1 YOECP	fimbrial protein p
3	43	100.0	188	1 YOECP	F7-2 fimbrial prot
4	40	93.0	182	2 I41204	A-minor fimbrial p
5	35	81.4	187	1 YOECP	KS71A fimbrial p
6	35	81.4	187	2 C43597	plin type F7-1 pr
7	35	81.4	187	2 C43597	hypothetical prote
8	35	81.4	187	2 C43597	hypothetical prote
9	34	79.1	134	2 T46114	hypothetical prote
10	34	79.1	136	2 T09477	hypothetical prote
11	34	79.1	136	2 T09477	hypothetical prote
12	33	76.7	195	2 B75017	probable oxidoredu
13	33	76.7	195	2 B75017	hypothetical prote
14	33	76.7	229	2 B90513	50S ribosomal prot
15	33	76.7	229	2 B90513	hypothetical prote
16	33	76.7	229	2 B90513	hypothetical prote
17	33	76.7	229	2 B90513	hypothetical prote
18	33	76.7	229	2 B90513	hypothetical prote
19	33	76.7	229	2 B90513	hypothetical prote
20	33	76.7	229	2 B90513	hypothetical prote
21	33	76.7	229	2 B90513	hypothetical prote
22	33	76.7	229	2 B90513	hypothetical prote
23	33	76.7	229	2 B90513	hypothetical prote
24	33	76.7	229	2 B90513	hypothetical prote
25	33	76.7	229	2 B90513	hypothetical prote
26	33	76.7	229	2 B90513	hypothetical prote
27	33	76.7	229	2 B90513	hypothetical prote
28	33	76.7	229	2 B90513	hypothetical prote
29	33	76.7	229	2 B90513	hypothetical prote

30	32	74.4	894	1 FAHUA2	alpha-actinin 2 -
31	32	74.4	897	2 S02032	alpha-actinin 2, s
32	32	74.4	899	2 T11578	probable lipoxigen
33	32	74.4	1749	2 S75071	hypothetical prote
34	32	74.4	1856	2 C95008	immunoglobulin A1
35	32	74.4	2938	2 T30249	cell proliferation
36	31	72.1	38	2 T11763	acetyl-CoA carboxy
37	31	72.1	172	2 G70234	conserved hypotnet
38	31	72.1	174	1 A31096	fimbrial protein p
39	31	72.1	180	2 A05072	RNA-directed DNA p
40	31	72.1	205	2 C89814	hypothetical prote
41	31	72.1	216	2 T03637	grp-binding protei
42	31	72.1	249	2 A87520	hypothetical prote
43	31	72.1	255	2 C71169	probable flagellin
44	31	72.1	299	2 F95244	UMP-glucose-1-phos
45	31	72.1	299	2 D98109	UMP-glucose-1-phos

#### ALIGNMENTS

RESULT 1  
S25206  
prsa protein - Escherichia coli (strain 1442) (fragment)  
C:Species: Escherichia coli  
C:Date: 17-Feb-1998 #sequence\_revision 26-Feb-1998 #text\_change 26-Aug-1999  
C:Accession: S25206  
R:Marklund, B.I.; Tennent, J.M.; Garcia, E.; Hamers, A.; Baga, M.; Lindberg, F.; Gaas  
Mol. Microbiol. 6, 2225-2242, 1992  
A:Title: Horizontal gene transfer of the Escherichia coli pap and prs pil operons as  
A:Reference number: S25205; MUID:93023852  
A:Accession: S25206  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-161 <MA2>  
A:Cross-references: EMBL:X62157; NID:942515; PIDN:CAA44083.1; PID:942516  
A:Experimental source: strain 1442  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, September 1992  
A:Genetics:  
A:Gene: prsa  
C:Superfamily: F7-2 fimbrial protein

Query Match  
Best Local Similarity 100.0%; Score 43; DB 2; Length 161;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PQGGGKVT 8  
|||||||  
DB 6 PQGGGKVT 13

RESULT 2  
YOECP  
fimbrial protein pap precursor - Escherichia coli  
N:Alternate names: pap pilI  
C:Species: Escherichia coli  
C:Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 16-Jul-1999  
C:Accession: A23221; S25216; A05229; S16395  
R:Baga, M.; Normark, S.; Hardy, J.; O'Hanley, P.; Larik, D.; Olsson, O.; Schoolnik, G.  
J. Bacteriol. 157, 330-333, 1984  
A:Title: Nucleotide sequence of the papA gene encoding the pap pilus subunit of human  
A:Accession: A23221  
A:Molecule type: DNA  
A:Residues: 1-185 <BAG>  
A:Cross-references: GB:X03391; GB:X03392; NID:942309; PIDN:CAA27126.1; PID  
R:Marklund, B.I.; Tennent, J.M.; Garcia, E.; Hamers, A.; Baga, M.; Lindberg, F.; Gaas  
Mol. Microbiol. 6, 2225-2242, 1992  
A:Title: Horizontal gene transfer of the Escherichia coli pap and prs pil operons as  
A:Reference number: S25205; MUID:93023852  
A:Accession: S25216  
A:Molecule type: DNA  
A:Residues: 1-185 <MA2>

A:Cross-references: EMBL:X61239; NID:g42290; PIDN:CAA43562.1; PID:g42293  
A:Experimental source: strain J96

C:Genetics:

A:Gene: papa

C:Superfamily: F7-2 fimbrial protein

C:Keywords: fimbria

F1:1-22/Domain: signal sequence #status predicted <SIG>

F1:23-185/Product: fimbrial protein papa #status predicted <MAT>

Query Match 100.0%; Score 43; DB 1; Length 185;

Best Local Similarity 100.0%; Pred. No. 0.51;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PQGGGKVT 8

|||||

Db 27 PQGGGKVT 34

### RESULT 3

F2 fimbrial protein precursor - Escherichia coli

N:Alternate names: F7-2 pilin

C:Species: Escherichia coli

C>Date: 28-Feb-1986 #sequence-revision 28-Feb-1986 #text-change 16-Jul-1999

C:Accession: A03496; B43597

R:Van Die, I.; Bergmans, H.

Gene 32, 83-90, 1984

A:Title: Nucleotide sequence of the gene encoding the F7-2 fimbrial subunit of a uropathogenic

A:Reference number: A03496; MUID:85155485

A:Accession: A03496

A:Residues: 1-188 <VAN>

A:Molecule type: DNA

A:Cross-references: GB:M12861; NID:g145963; PIDN:AAA23778.1; PID:g145964

R:Denich, K.; Blyn, L.B.; Crail, A.; Braaten, B.A.; Hardy, J.; Low, D.A.; O'Hanley, P.D.

Infect. Immun. 59, 3849-3858, 1991

A:Title: DNA sequences of three papa genes from uropathogenic Escherichia coli strains:

A:Reference number: A43597; MUID:92040048

A:Accession: B43597

A:Molecule type: DNA

A:Residues: 1-188 <DEN>

A:Cross-references: GB:M68060; NID:g147070; PIDN:AAA24278.1; PID:g147071

C:Genetics:

A:Gene: papa

C:Function: A:Description: one of the fimbrial proteins involved in mannose-resistant hemagglutination

subunits into fimbriae, or regulation of phase variation

C:Superfamily: F7-2 fimbrial protein

C:Keywords: fimbria; mannose-resistant hemagglutination

F1:21/Domain: signal sequence #status predicted <SIG>

F1:22-188/Product: F7-2 fimbrial protein #status predicted <MAT>

Query Match 100.0%; Score 43; DB 1; Length 188;

Best Local Similarity 100.0%; Pred. No. 0.51; Mismatches 0; Indels 0; Gaps 0;

OY 1 PQGGGKVT 8

|||||

Db 26 PQGGGKVT 33

### RESULT 4

A:minor fimbrial protein - Escherichia coli

C:Species: Escherichia coli

C>Date: 02-Jul-1996 #sequence-revision 02-Jul-1996 #text-change 21-Jul-2000

C:Accession: I41204; A44853

R:Mail, S.N.; Desgroselliers, L.; Fairbrother, J.M.; Harel, J.

Microb. Pathog. 16, 15-25, 1994

A:Title: Analysis of genes coding for the major and minor fimbrial subunits of the Prs-1

A:Reference number: I41201; MUID:94335605

A:Accession: I41204

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-182 <RES>

A:Cross-references: GB:I07420; NID:g1228096; PIDN:AAA92574.1; PID:g452584

R:Harel, J.; Forget, C.; Saint-Amant, J.; Daigle, F.; Dubreuil, D.; Jacques, M.; Fair

J. Gen. Microbiol. 138, 1495-1502, 1992

A:Title: Molecular cloning of a determinant coding for fimbrial antigen Fl65(1), a Pr

A:Contents: A44853; MUID:92381496

A:Accession: A44853

A:Status: preliminary

A:Molecule type: protein

A:Residues: 22-30, 1, 32-35 <HAR>

A>Note: sequence extracted from NCBI backbone (NCBIP:112336)

C:Superfamily: F7-2 fimbrial protein

Query Match 93.0%; Score 40; DB 2; Length 182;

Best Local Similarity 87.5%; Pred. No. 1.9; Mismatches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 PQGGGKVT 8

|||||

Db 26 PQGGGKVT 33

### RESULT 5

KS71A fimbrial protein precursor - Escherichia coli

N:Alternate names: KS71A pilin

C:Species: Escherichia coli

C>Date: 31-Mar-1988 #sequence-revision 31-Mar-1988 #text-change 16-Jul-1999

C:Accession: A23117

R:Rhen, M.; van Die, I.; Rhen, V.; Bergmans, H.

Eur. J. Biochem. 151, 573-577, 1985

A:Title: Comparison of the nucleotide sequences of the genes encoding the KS71A and F

A:Reference number: A23117; MUID:85285072

A:Accession: A23117

A:Molecule type: DNA

A:Residues: 1-187 <RHE>

A:Cross-references: GB:X02921; NID:g41880; PIDN:CAA26678.1; PID:g41881

C:Superfamily: F7-2 fimbrial protein

C:Keywords: fimbria

F1:21/Domain: signal sequence #status predicted <SIG>

F1:22-187/Product: KS71A fimbrial protein #status predicted <MAT>

Query Match 81.4%; Score 35; DB 1; Length 187;

Best Local Similarity 75.0%; Pred. No. 19; Mismatches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 PQGGGKVT 8

|||||

Db 26 PQGGGKVT 33

### RESULT 6

pilin type F7-1 precursor - Escherichia coli

C:Species: Escherichia coli

C>Date: 12-Jan-1993 #sequence-revision 12-Jan-1993 #text-change 26-Aug-1999

C:Accession: C43597

R:Denich, K.; Blyn, L.B.; Crail, A.; Braaten, B.A.; Hardy, J.; Low, D.A.; O'Hanley, P

Infect. Immun. 59, 3849-3858, 1991

A:Title: DNA sequences of three papa genes from uropathogenic Escherichia coli strain

A:Reference number: A43597; MUID:92040048

A:Accession: C43597

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-187 <DEN>

A:Cross-references: GB:M68061; NID:g147072; PIDN:AAA24279.1; PID:g147073

C:Superfamily: F7-2 fimbrial protein

Query Match 81.4%; Score 35; DB 2; Length 187;

Best Local Similarity 75.0%; Pred. No. 19; Mismatches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
|||||:|:  
DB 26 POGGKVS 33

## RESULT 7

576891

hypothetical protein sl10443 - *Synechocystis* sp. (strain PCC 6803)C:Species: *Synechocystis* sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000

C:Accession: T76891

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*

5.

A:Reference number: S74322; MUID:97061201

A:Accession: S76891

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-469 &lt;KAN&gt;

A:Cross-references: EMBL:D90917; GB:AB001339; NID:g1653836; PID:BAA1803.1; PID:g165389

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Start codon: GTG

C:Superfamily: *Synechocystis* hypothetical protein sl10443

Query Match 81.4%; Score 35; DB 2; Length 469;  
Best Local Similarity 85.7%; Pred. No. 46;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 7  
|||||:  
DB 403 POGGKRL 409

## RESULT 8

T42671

hypothetical protein DKFZp434I1117.1 - human (fragment)

C:Species: *Homo sapiens* (man)

C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 11-Jan-2000

C:Accession: T42671

R:Blöcker, H.; Boecher, M.; Brandt, P.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, November 1999

A:Reference number: Z22230

A:Accession: T42671

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-613 &lt;AAA&gt;

A:Cross-references: EMBL:AL133071

A:Experimental source: adult testis; clone DKFZp434I1117

C:Genetics:

A:Note: DKFZp434I1117.1

Query Match 81.4%; Score 35; DB 2; Length 613;  
Best Local Similarity 85.7%; Pred. No. 60;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 7  
|||||:  
DB 417 POGGKRV 423

## RESULT 9

T46114

hypothetical protein T2783.70 - *Arabidopsis thaliana*C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 04-Feb-2000 #sequence\_revision 04-Feb-2000 #text\_change 02-Sep-2000

C:Accession: T46114

R:Nyakatura, G.; Fartmann, B.; Dauner, D.; Steier, W.; Holland, R.; Weichselgartner, M.;

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23022

A:Accession: T46114  
A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-134 &lt;NTA&gt;

A:Cross-references: EMBL:AL137079

A:Experimental source: cultivar Columbia; BAC clone T27B3

C:Genetics:

A:Map position: 3

A:Insertion: 58/3

A:Note: T27B3.70

C:Superfamily: *Arabidopsis thaliana* hypothetical protein T21C14.30

Query Match 79.1%; Score 34; DB 2; Length 134;  
Best Local Similarity 85.7%; Pred. No. 21;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 7  
|||||:  
DB 14 POGGKRV 20

## RESULT 10

T09477

hypothetical protein - human

C:Species: *Homo sapiens* (man)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 16-Jul-1999

C:Accession: T09477

R:Murphy, L.

submitted to the EMBL Data Library, January 1998

A:Reference number: Z16683

A:Accession: T09477

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-136 &lt;MUR&gt;

A:Cross-references: EMBL:AL021546; NID:e1248287; PID:e1248290

C:Genetics:

A:Gene: 15E1.2

A:Map position: 12

A:Insertion: 27/3; 85/2; 120/1

Query Match 79.1%; Score 34; DB 2; Length 136;  
Best Local Similarity 62.5%; Pred. No. 21;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
|||||:  
DB 26 POGGKRV 33

## RESULT 11

B70017

probable oxidoreductase (EC 1.-.-.-) yurc [similarity] - *Bacillus subtilis*C:Species: *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 21-Jul-2000

C:Accession: B70017

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.F.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritze, C.; Fujita, M.; Fujita, Y.; Funai, S.; Gallizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauder, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rochoa, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestre, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipac, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbach, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*

A:Reference number: A69580; MUID:98044033

A:Accession: B70017

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA  
A:Residues: 1-745 <KUN>  
A:Cross-references: GB:299120; GB:AL009126; NID:g2635613; PIDN:CAB15238.1; PID:g2635745  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: yurC  
C:Superfamily: carbon-monoxide dehydrogenase large chain  
C:Keywords: oxidoreductase

Query Match 79.1%; Score 34; DB 2; Length 745;  
Best Local Similarity 75.0%; Pred. NO. 1.1e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
| | | | |  
Db 11 POGGKVT 18

## RESULT 12

hthetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 26-May-2000

C:Accession: E15273  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.D.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Mc  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896

A:Accession: E15273

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-195 <WHI>

A:Cross-references: GB:AE002074; GB:AE000513; NID:g6460257; PIDN:AAE11995.1; PID:g646027

A:Experimental source: strain R1

C:Genetics:

A:Gene: DR2446

A:Map position: 1

C:Superfamily: Deinococcus radiodurans hypothetical protein DR2446

Query Match 76.7%; Score 33; DB 2; Length 195;  
Best Local Similarity 75.0%; Pred. NO. 48;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
| | | | |  
9 POGGKVT 16

## RESULT 13

50S ribosomal protein L1 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 03-Aug-2001

C:Accession: B90513

R:Chambaud, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallisson, F.; Moszer, I.;  
Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma pulm

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: B90513

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-229 <KUR>

A:Cross-references: GB:AL445566; PID:g14089423; PIDN:CAC13183.1; GSPDB:GN00153

A:Experimental source: strain UAB CTIP

C:Genetics:

A:Gene: MYP0\_0100

A:Genetic code: SGC3

C:Superfamily: Escherichia coli ribosomal protein L1

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
| | | | |  
Db 66 POGGKVT 73

## RESULT 14

hypothetical protein RV2812 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 22-Oct-1999

C:Accession: G70690

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon  
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,  
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987

A:Accession: G70690

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-469 <COL>

A:Cross-references: GB:281331; GB:AL123456; NID:g3261650; PIDN:CAB03635.1; PID:e27680

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2812

Query Match 76.7%; Score 33; DB 2; Length 469;  
Best Local Similarity 71.4%; Pred. NO. 1.1e+02;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 7  
| | | | |  
Db 256 POGGKVT 262

## RESULT 15

Probable transposase NMA1882 [imported] - Neisseria meningitidis (strain Z2491 serogr

C:Species: Neisseria meningitidis

C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001

C:Accession: E81815

R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491

A:Reference number: A81775; MUID:20222556

A:Accession: E81815

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-681 <PAR>

A:Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85105.1; PID:g738

A:Experimental source: serogroup A, strain Z2491

C:Genetics:

A:Gene: NMA1882

Query Match 76.7%; Score 33; DB 2; Length 681;  
Best Local Similarity 71.4%; Pred. NO. 1.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 7  
| | | | |  
Db 398 POGGKVT 404

Search completed: October 28, 2002, 17:23:13  
Job time : 13.7895 secs



```

ID PAPA_ECOLI STANDARD; PRT; 185 AA.
AC P04127;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE PAP fimbrial major pilin protein precursor (PAP pilin).
GN PAPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=J96;
RX MEDLINE=84087728; PubMed=6140260;
RA Baga M., Normark S., Hardy J., O'Hanley P., Lark D., Olsson O.,
RT Scholnik G., Falkow S.;
RT "Nucleotide sequence of the papA gene encoding the Pap pilus subunit
of human uropathogenic Escherichia coli.";
RL J. Bacteriol. 157:330-333(1984).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=J96;
RX MEDLINE=93023852; PubMed=1357526;
RA Marklund B.-T., Tennent J.M., Garcia E., Hamers A., Baga M.,
RA Lindberg F., Gastra W., Normark S.;
RT "Horizontal gene transfer of the Escherichia coli pap and prs pil1
operons as a mechanism for the development of tissue-specific
adhesive properties.";
RL Mol. Microbiol. 6:2225-2242(1992).
CC -1- FUNCTION: FIMBRIAL (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DISEASE: STRAINS OF E. COLI THAT CAUSE INFECTION OF THE HUMAN
URINARY TRACT PRODUCE PAP-PILI WHICH ARE HAIR-LIKE APPENDAGES
CONSISTING OF ABOUT 1000 HELICALLY ARRANGED SUBUNITS OF THE
PROTEIN PAPA. THESE PILI MEDIATE BINDING TO DIGALACTOSIDE-
CONTAINING GLYCOLIPIDS PRESENT ON THE EPITHELIAL CELLS WHICH
LINE THE URINARY TRACT.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: X03391; CAA27126.1; -
CC EMBL: X61239; CAA43562.1; -
DR PIR: A23221; YOECP.
DR InterPro: IPR000259; Fimbrlal.
DR Pfam: PF00419; Fimbrlal; 1.
KW Fimbrlal; Signal.
FT SIGNAL 1 22
FT CHAIN 23 185
FT DISUFID 44 83
SO SEQUENCE 185 AA; 18686 MW; 93DB4FFDA211CG71 CRC64;

Query Match 100.0%; Score 43; DB 1; Length 185;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 POGGKVT 8
Db 27 POGGKVT 34

RESULT 3
FME2_ECOLI

```

```

ID FME2_ECOLI STANDARD; PRT; 188 AA.
AC P02972;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE F7-2 fimbrial protein precursor (F7-2 pilin).
GN F7-2 OR PAPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN (1)
RP SEQUENCE FROM N.A.
RC MEDLINE=85155489; PubMed=6152241;
RA van Die I., Bergmans H.;
RT "Nucleotide sequence of the gene encoding the F72 fimbrial subunit of
a uropathogenic Escherichia coli strain.";
RL Gene 32:83-90(1984).
[2]
RP SEQUENCE FROM N.A.
RC MEDLINE=92040048; PubMed=1682251;
RA Denich K., Blyn L.B., Craiu A., Braaten B.A., Hardy J., Low D.A.,
RA O'Hanley P.D.;
RT "DNA sequences of three papA genes from uropathogenic Escherichia
coli strains: evidence of structural and serological conservation.";
RL Infect. Immun. 59:3849-3858(1991).
CC -1- FUNCTION: FIMBRIAL (ALSO CALLED PILI), POLAR FILAMENTS RADIATING
FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5
MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO
COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.
CC -1- DISEASE: THIS IS ONE OF THE FIMBRIAL PROTEINS INVOLVED IN
MANNOSE-RESISTANT HEMAGGLUTINATION OF HUMAN ERYTHROCYTES.
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: M12861; AAA23778.1; -
CC EMBL: M68060; AAA24278.1; -
DR PIR: A03496; YOECP2.
DR PIR: B43597; B43597.
DR InterPro: IPR000259; Fimbrlal.
DR Pfam: PF00419; Fimbrlal; 1.
KW Fimbrlal; Signal.
FT SIGNAL 1 21
FT CHAIN 22 188
FT DISUFID 43 82
SO SEQUENCE 188 AA; 19184 MW; 0EEF750CFB843157 CRC64;

Query Match 100.0%; Score 43; DB 1; Length 188;
Best Local Similarity 100.0%; Pred. No. 0.19;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 POGGKVT 8
Db 26 POGGKVT 33

RESULT 4
FME2_ECOLI
ID FME2_ECOLI STANDARD; PRT; 187 AA.
AC P04740;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE KS71A fimbriillin precursor (P-fimbrial antigen).
GN KS71A.
OS Escherichia coli.

```

CC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
CC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=65285072; PubMed=2992970;  
RA Rhen M., van Die I., Rhen V., Bergmans H.;  
RT "Comparison of the nucleotide sequences of the genes encoding the  
RT KS71A and F7(1) fimbrial antigens of uropathogenic Escherichia  
RT coli";  
RL Eur. J. Biochem. 151:573-577(1995).  
CC -1- FUNCTION: FIMBRIAE (ALSO CALLED PILI), POLAR FILAMENTS RADIATING  
CC FROM THE SURFACE OF THE BACTERIUM TO A LENGTH OF 0.5-1.5  
CC MICROMETERS AND NUMBERING 100-300 PER CELL, ENABLE BACTERIA TO  
CC COLONIZE THE EPITHELIUM OF SPECIFIC HOST ORGANS.  
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X02921; CAA26678.1; -  
DR PIR; A23117; YOECKS.  
DR InterPro: IPR000259; Fimbr1al.  
DR Pfam; PF00419; Fimbr1al; 1.  
KW Fimbria; Signal; Antigen.  
FT SIGNAL 1 21 POTENTIAL.  
FT CHAIN 22 187 KS71A FIMBRILIN.  
FT DISULFID 43 82 PROBABLE.  
SQ SEQUENCE 187 AA; 19310 MW; 799FA38264C638DC CRC64;  
  
Query Match 81.4%; Score 35; DB 1; Length 187;  
Best Local Similarity 75.0%; Pred. No. 7;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
  
OY 1 POGGSKVT 8  
DB 26 POGGSEVS 33  
  
RESULT 5  
1SE2\_HUMAN STANDARD; PRT; 136 AA.  
AC 045716;  
CC 30-MAY-2000 (Rel. 39, Created)  
CC 30-MAY-2000 (Rel. 39, Last sequence update)  
CC 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Putative protein 1SE1.2.  
GN 15L1.2.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Murphy L.;  
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: BELONGS TO THE GATC FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AL021546; CAA16496.1; -  
DR InterPro: IPR003837; Glu-TRNAGln.  
DR

DR Pfam; PF02686; Glu-TRNAGln; 1.  
KW Hypothetical protein.  
SQ SEQUENCE 136 AA; 15086 MW; 113118E9507234E4 CRC64;  
  
Query Match 79.1%; Score 34; DB 1; Length 136;  
Best Local Similarity 62.5%; Pred. No. 8.2;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 POGGSKVT 8  
DB 26 POGGSRIT 33  
  
RESULT 6  
RUN2\_HORSE STANDARD; PRT; 131 AA.  
AC 09XSB7;  
DT 01-MAR-2002 (Rel. 41, Created)  
DT 01-MAR-2002 (Rel. 41, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Runt-related transcription factor 2 (Core-binding factor, alpha 1  
DE subunit) (CBF-alpha 1) (Fragment).  
GN RUNX2 OR CBFAL.  
OS Equus caballus (Horse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.  
OX NCBI\_TaxID=9796;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Uemura T., Fu Z.W., Kato H.;  
RT "Molecular cloning and expression of equine Cbfal.";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: Transcription factor involved in osteoblastic  
CC differentiation and skeletal morphogenesis. Essential for the  
CC maturation of osteoblasts and both intramembranous and  
CC endochondral ossification. Cbf binds to the core site, 5'-  
CC PGGYGGT-3', of a number of enhancers and promoters, including  
CC murine leukemia virus, polyomavirus enhancer, T-cell receptor  
CC enhancers, osteocalcin, osteopontin, bone sialoprotein, alpha 1(I)  
CC collagen, Lck, IL-3 and GM-CSF promoters (By similarity).  
CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha  
CC subunit binds DNA as a monomer and through the runt domain. DNA-  
CC binding is increased by heterodimerization (By similarity).  
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -1- PTM: Phosphorylated; probably by MAP kinases (MAPK) (By  
CC similarity).  
CC -1- SIMILARITY: CONTAINS 1 RUNT DOMAIN.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF113507; AAD26154.1; -  
KW Transcription regulation; DNA-binding; Nuclear protein;  
KW Multigene family; Phosphorylation.  
FT NON\_TER 1 1  
FT DOMAIN <1 10 RUNT.  
FT NON\_TER 131 131  
SQ SEQUENCE 131 AA; 14249 MW; 2516955230654247 CRC64;  
  
Query Match 74.4%; Score 32; DB 1; Length 131;  
Best Local Similarity 62.5%; Pred. No. 19;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;  
  
OY 1 POGGSKVT 8  
DB 23 POGGSOIT 30

```

RESULT 7
IDHP_MEDSA          STANDARD;          PRT;          433 AA.
AC 040345;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Isolate dehydrogenase [NADP], chloroplast precursor (EC 1.1.1.42)
DE (Oxalosuccinate decarboxylase) (IDH) (NADP+-specific IDDH) (IDP)
DE (Fragment).
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae; Medicago.
OX NCBI_TaxID=3879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93099232; PubMed=1281435;
PT Shorrock B.S., Dixon R.A.;
DE Molecular characterization and expression of an isocitrate
DE dehydrogenase from alfalfa (Medicago sativa L.)";
RL Plant Mol. Biol. 20:801-807(1992).
CC -1- CATALYTIC ACTIVITY: Isocitrate + NADP(+) = 2-oxoglutarate + CO(2)
CC + NADPH.
CC -1- CORRECTOR: REQUIRES MANGANESE OR MAGNESIUM (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Chloroplast (Potential).
CC -1- TISSUE SPECIFICITY: DETECTED IN ALL TISSUES EXAMINED.
CC -1- SIMILARITY: BELONGS TO THE ISOCITRATE AND ISOPROPYLMALATE
CC DEHYDROGENASES FAMILY.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M3672; AAA32656.1; -
DR InterPro; IPR001804; Isoch.
DR Pfam; PF00180; Isoch; 1.
DR PROSITE; PS00470; IDH_IMDH; 1.
KW Oxidoreductase; NADP; glyoxylate bypass; Tricarboxylic acid cycle;
KW Transit peptide; Chloroplast.
FT TRANSIT 1 21 CHLOROPLAST (BY SIMILARITY).
FT CHAIN 22 433 ISOCITRATE DEHYDROGENASE [NADP].
FT ACT_SITE 117 117 BINDING TO ISOCITRATE (BY SIMILARITY).
FT SEQUENCE 433 AA; 46383 MW; E4421317B8A2CFE3 CRC64;
Query Match 74.4%; Score 32; DB 1; Length 433;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 PGGGKVT 8
I:|:|:|:|
DB 181 PEGGERT 188

```

```

RESULT 8
IDHP_MEDSA          STANDARD;          PRT;          521 AA.
AC 013950; 014614; 014615; 095181;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE RunT-related transcription factor 2 (Core-binding factor, alpha 1
DE subunit) (CBF-alpha 1) (Acute myeloid leukemia 3 protein) (Oncogene
DE AML-3) (Polymavirus enhancer binding protein 2 alpha A subunit)
DE (PEB2-alpha A) (PEA2-alpha A) (SL3-3 enhancer factor 1 alpha A
DE subunit) (SL3/KV core-binding factor alpha A subunit) (Osteoblast-
DE specific transcription factor 2) (OSF-2).
RN RUNX2 OR CBFAL OR AML3 OR PEBP2A OR OSF2.

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2). VARIANT CCD 75-ALA--ALA-
RP 84 DEL. AND VARIANT 78-ALA--ALA-83 DEL.
RX MEDLINE=97325753; PubMed=9182765;
RA Mundlos S., Otto F., Mundlos C., Mulliken J.B., Aylsworth A.S.,
RA Albrigth S., Lindhout D., Cole W.G., Henn W., Knoll J.H.M., Owen M.J.,
RA Merzlsman R., Zabel B.U., Olsen B.R.;
RT "Mutations involving the transcription factor CBFAL cause
RT cleidocranial dysplasia.";
RL Cell 89:773-779(1997).
RN [2]
RP SEQUENCE OF 60-521 FROM N.A. (ISOFORM 3).
RX MEDLINE=97377000; PubMed=9233771;
RA Zhang Y.-W., Bae S.-C., Takahashi E.-T., Ito Y.;
RT "The cDNA cloning of the transcripts of human PEBP2alpha/CBFAL mapped
RT to 6p12.3-p21.1, the locus for cleidocranial dysplasia.";
RL Oncogene 15:367-371(1997).
RN [3]
RP SEQUENCE OF 1-19 FROM N.A. (ISOFORM 1).
RX MEDLINE=9832266; PubMed=9651525;
RA Xiao Z.S., Thomas R., Hinson T.K., Quarles L.D.;
RT "Genomic structure and isoform expression of the mouse, rat and human
RT Cbfa1/Osf2 transcription factor.";
RL Gene 214:187-197(1998).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3). AND ALTERNATIVE SPLICING.
RX MEDLINE=98096586; PubMed=9434946;
RA Geoffroy V., Corral D.A., Zhou L., Lee B., Karsenty G.;
RT "Genomic organization, expression of the human CBFAL gene, and
RT evidence for an alternative splicing event affecting protein
RT function.";
RL Mamm. Genome 9:54-57(1998).
RN [5]
RP VARIANTS CCD ARG-175 AND ASN-191.
RX MEDLINE=97351518; PubMed=9207800;
RA Lee B., Thirunavukkarasu K., Zhou L., Pastore L., Baldini A.,
RA Hecht J., Geoffroy V., Ducey P., Karsenty G.;
RT "Missense mutations abolishing DNA binding of the osteoblast-specific
RT transcription factor Osf2/CBFAL in cleidocranial dysplasia.";
RL Nat. Genet. 16:307-310(1997).
RN [6]
RP VARIANTS CCD R-113; R-118; C-121; R-123; R-205; Q-225; W-225 AND
RP S-511.
RX MEDLINE=99452588; PubMed=10521292;
RA Quack I., Vonderstrass B., Stock M., Aylsworth A.S., Becker A.,
RA Brunton L., Lee P.J., Majewski F., Mulliken J.B., Suri M., Zenker M.,
RA Mundlos S., Otto F.;
RT "Mutation analysis of core binding factor A1 in patients with
RT cleidocranial dysplasia.";
RL Am. J. Hum. Genet. 65:1268-1278(1999).
RN [7]
RP VARIANTS CCD N-133 DEL; Q-169; R-175; Q-190; N-191; C-193; F-199;
RP A-200; R-209 AND Q-225.
RX MEDLINE=20014723; PubMed=10545612;
RA Zhou G., Chen Y., Zhou L., Thirunavukkarasu K., Hecht J., Chitayat D.,
RA Gelb B.D., Pirinen S., Berry S.A., Greenberg C.R., Karsenty G.,
RA Lee B.;
RT "CBFAL mutation analysis and functional correlation with phenotypic
RT variability in cleidocranial dysplasia.";
RL Hum. Mol. Genet. 8:2311-2316(1999).
RN [8]
RP VARIANT CCD SER-197.
RX MEDLINE=20156388; PubMed=10689183;
RA Zhang Y.-W., Yasui N., Kakazu N., Abe T., Takada K., Imai S., Sato M.,
RA Nomura S., Ochi T., Okuzumi S., Nogami H., Nagai T., Ohashi H.,
RA Ito Y.;
RT "PEBP2alpha/CBFAL mutations in Japanese cleidocranial dysplasia
RT patients.";
RL Gene 244:21-28(2000).

```



RN [9]  
 RP VARIANT CCD TRP-190.  
 RX MEDLINE=20438365; PubMed=10980549;  
 RA Giannotti A., Tessa A., Patrono C., De Florio L., Velardo M.,  
 RA Dionisi-Vici C., Bertini E., Santorelli F.M.;  
 RT "A novel CBFA1 mutation (R190W) in an Italian family with  
 RT cleidocranial dysplasia.";  
 RL Hum. Mutat. 16:277-277(2000).  
 CC -1- FUNCTION: Transcription factor involved in osteoblastic  
 CC differentiation and skeletal morphogenesis. Essential for the  
 CC maturation of osteoblasts and both intramembranous and  
 CC endochondral ossification. Cbf1 binds to the core site, 5'-  
 CC PYPGCGT-3', of a number of enhancers and promoters, including  
 CC murine leukemia virus, polyomavirus enhancer, T-cell receptor  
 CC enhancers, osteocalcin, osteopontin, bone sialoprotein, alpha 1(I)  
 CC collagen, Lck, IL-3 and GM-CSF promoters (By similarity).  
 CC -1- SUBUNIT: Heterodimer of an alpha and a beta subunit. The alpha  
 CC subunit binds DNA as a monomer and through the Runt domain. DNA-  
 CC binding is increased by heterodimerization.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 3 isoforms; 1/Cbfa1 (shown here), 2 and  
 CC 3/Cbfa1b; are produced by alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Specifically expressed in osteoblasts.  
 CC -1- DOMAIN: A proline/serine/threonine rich region at the C-terminus  
 CC is necessary for transcriptional activation of target genes and  
 CC contains the phosphorylation sites.  
 CC -1- PTM: Phosphorylated; probably by MAP kinases (MAPK) (By  
 CC similarity).  
 CC -1- DISEASE: Defects in RUNX2 are the cause of cleidocranial dysplasia  
 CC (CCD), an autosomal dominant skeletal disorder with high  
 CC penetrance and variable expressivity. It is due to defective  
 CC endochondral and intramembranous bone formation. Typical features  
 CC include hypoplasia/aplasia of clavicles, patent fontanelles,  
 CC wormian bones (additional cranial plates caused by abnormal  
 CC ossification of the calvaria), supernumerary teeth, short stature,  
 CC and other skeletal changes. In some cases defects in RUNX2 are  
 CC exclusively associated with dental anomalies.  
 CC -1- SIMILARITY: CONTAINS 1 RUNT DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; AF001450; AAB65159.2; -  
 DR EMBL; AF001443; AAB65159.2; JOINED.  
 DR EMBL; AF001444; AAB65159.2; JOINED.  
 DR EMBL; AF001445; AAB65159.2; JOINED.  
 DR EMBL; AF001446; AAB65159.2; JOINED.  
 DR EMBL; AF001447; AAB65159.2; JOINED.  
 DR EMBL; AF001448; AAB65159.2; JOINED.  
 DR EMBL; AF001449; AAB65159.2; JOINED.  
 DR EMBL; AF001450; AAB65158.1; -  
 DR EMBL; AF001444; AAB65158.1; JOINED.  
 DR EMBL; AF001445; AAB65158.1; JOINED.  
 DR EMBL; AF001446; AAB65158.1; JOINED.  
 DR EMBL; AF001447; AAB65158.1; JOINED.  
 DR EMBL; AF001448; AAB65158.1; JOINED.  
 DR EMBL; AF001449; AAB65158.1; JOINED.  
 DR EMBL; AF001450; AAB65158.1; JOINED.  
 DR EMBL; AF003952; AAC78624.1; -  
 DR EMBL; AF003949; AAC77441.1; -  
 DR MIM; 600211; -  
 DR MIM; 119600; -  
 DR InterPro; IPR000040; AML1\_Runt.  
 DR Pfam; PF00853; Runt; 1.  
 DR PRINTS; PR00967; ONCOGENEAML1.  
 KW Transcription regulation; DNA-binding; Nuclear protein; ATP-binding;  
 KW Alternative splicing; Phosphorylation; Polymorphism; Disease mutation.  
 FT DOMAIN 101 228 RUNT.

FT	DOMAIN	237	521	PRO/SER/THR-RICH.
FT	DOMAIN	49	71	POLY-GLN.
FT	DOMAIN	73	89	POLY-ALA.
FT	NP_BIND	189	196	ATP (POTENTIAL).
FT	VARSPPLIC	1	19	MASNSLSTVPCQONFEW -> MRIPV (IN ISOFORM 2).
FT	VARSPPLIC	341	362	MISSING (IN ISOFORM 3).
FT	VARSPPLIC	75	84	AAAAAAAAA -> AAAAAAAAAAAAAAAAAA (IN CCD ASSOCIATED WITH BRACHYDACTYL OF HANDS AND FEET).
FT	VARSPPLIC	78	83	MISSING.
FT	VARSPPLIC	113	113	/FTID=VAR_012131.
FT	VARSPPLIC	118	118	L -> R (IN CCD).
FT	VARSPPLIC	121	121	/FTID=VAR_012132.
FT	VARSPPLIC	121	121	S -> R (IN CCD).
FT	VARSPPLIC	123	123	/FTID=VAR_012133.
FT	VARSPPLIC	123	123	F -> C (IN CCD).
FT	VARSPPLIC	133	133	/FTID=VAR_012134.
FT	VARSPPLIC	133	133	C -> R (IN CCD).
FT	VARSPPLIC	133	133	/FTID=VAR_012135.
FT	VARSPPLIC	133	133	MISSING (IN CCD).
FT	VARSPPLIC	169	169	/FTID=VAR_012136.
FT	VARSPPLIC	175	175	R -> Q (IN CCD).
FT	VARSPPLIC	190	190	/FTID=VAR_012137.
FT	VARSPPLIC	190	190	M -> R (IN CCD).
FT	VARSPPLIC	190	190	/FTID=VAR_012138.
FT	VARSPPLIC	190	190	R -> Q (IN CCD).
FT	VARSPPLIC	190	190	/FTID=VAR_012139.
FT	VARSPPLIC	190	190	R -> W (IN CCD).
FT	VARSPPLIC	191	191	S -> N (IN CCD).
FT	VARSPPLIC	191	191	/FTID=VAR_012140.
FT	VARSPPLIC	193	193	R -> C (IN CCD).
FT	VARSPPLIC	193	193	/FTID=VAR_012141.
FT	VARSPPLIC	197	197	R -> S (IN CCD).
FT	VARSPPLIC	197	197	/FTID=VAR_012142.
FT	VARSPPLIC	199	199	F -> S (IN CCD).
FT	VARSPPLIC	199	199	/FTID=VAR_012143.
FT	VARSPPLIC	200	200	L -> F (IN CCD).
FT	VARSPPLIC	200	200	/FTID=VAR_012144.
FT	VARSPPLIC	200	200	T -> A (IN CCD).
FT	VARSPPLIC	200	200	WITH ISOLATED DENTAL ANOMALIES; NORMAL DNA BINDING).

Query Match 74.4%; Score 32; DB 1; Length 521;  
 Best Local Similarity 62.5%; Pred. No. 72;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 PGGGQKVT 8  
 DB 279 PGGGQKVT 286

RESULT 9  
 AC RUN2\_MOUSE STANDARD; PRT; 607 AA.  
 ID 008775; 008776; 035183; 090217; 090U06; 090U04; 090Y29; 090JN0;  
 DT 01-MAR-2002 (Rel. 41, Created)  
 DT 01-MAR-2002 (Rel. 41, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Runt-related transcription factor 2 (Core-binding factor, alpha 1  
 DE subunit) (CBF-alpha 1) (Acute myeloid leukemia 3 protein) (Oncogene  
 DE AM-3) (Polyomavirus enhancer binding protein 2 alpha 1 subunit)  
 DE (PEBP2-alpha A) (PEA2-alpha A) (SL3-3 enhancer factor 1 alpha A  
 DE subunit) (SL3/ARY core-binding factor alpha A subunit) (Osteoblast-  
 DE specific transcription factor 2) (OSF-2).  
 GN RUNX2 OR CBFA1 OR AML3 OR PEBP2A OR OSF2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP MEDLINE=93342088; PubMed=8341710;  
 RX SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).

RA Ogawa E., Maruyama M., Kagoshima H., Inuzuka M., Lu J., Satake M.,  
RA Shigesada K., Ito Y. ;  
RT "PEBP2/PEA2 represents a family of transcription factors homologous to  
RT the products of the Drosophila runt gene and the human AML1 gene." ;  
RL Proc. Natl. Acad. Sci. U.S.A. 90:6859-6863(1993).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC STRAIN-C57BL/6 ;  
RC TISSUE-Osteoblast ;  
RX MEDLINE-97332570 ; PubMed-9182762 ;  
RA Ducky P., Zhang R., Geoffroy V., Ridall A.L., Karsenty G. ;  
RT "Ostf/Cbfa1: a transcriptional activator of osteoblast  
RT differentiation." ;  
RL Cell 89:747-754(1997).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORMS 2, 3, 4, 5, 6, 7, 8 AND 9).  
RC STRAIN-CD2-MYC ;  
RC MEDLINE-97385157 ; PubMed-9238031 ;  
RX Stewart M., Terry A., Hu M., O'Hara M., Blyth K., Baxter E.,  
RA Cameron E., Onions D.E., Nell J.C. ;  
RT "Proximal insertions induce the expression of bone-specific isoforms  
RT of PEBP2alpha (CBFA1): evidence for a new myc collaborating  
RL oncogene." ;  
RN Proc. Natl. Acad. Sci. U.S.A. 94:8646-8651(1997).  
RN [4]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2 AND 6), AND ALTERNATIVE  
RP SPLICING.  
RX MEDLINE-98322266 ; PubMed-9651525 ;  
RA Xiao Z.S., Thomas R., Hinson T.K., Quarles L.D. ;  
RT "Genomic structure and isoform expression of the mouse, rat and human  
RT Cbfa1/Ostf2 transcription factor." ;  
RL Gene 214:187-197(1998).  
RN [5]  
RP SEQUENCE OF 1-98 FROM N.A. (ISOFORMS 1 AND 2).  
RX MEDLINE-99453726 ; PubMed-10524201 ;  
RA Fujiwara M., Tagashira S., Harada H., Ogawa S., Katsumata T.,  
RA Nakatsuka M., Komori T., Takada H. ;  
RT "Isolation and characterization of the distal promoter region of mouse  
RT Cbfa1." ;  
RL Biochim. Biophys. Acta 1446:265-272(1999).  
RN [6]  
RP SEQUENCE OF 263-277 AND 305-319.  
RX MEDLINE-93242761 ; PubMed-8386878 ;  
RA Ogawa E., Inuzuka M., Maruyama M., Satake M., Naito-Fujimoto M.,  
RA Ito Y., Shigesada K. ;  
RT "Molecular cloning and characterization of PEBP2 beta, the  
RT heterodimeric partner of a novel Drosophila runt-related DNA binding  
RL protein PEBP2 alpha." ;  
RN Virology 194:314-331(1993).  
RN [7]  
RP SEQUENCE OF 1-35 FROM N.A.  
RC STRAIN-129 ;  
RA Chl X.-Z., Bae S.-C. ;  
RT "Analysis of the two PEBP2alpha/cbfa1 promoter regions." ;  
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.  
RN [8]  
RP FUNCTION.  
RX MEDLINE-97325751 ; PubMed-9182763 ;  
RA Komori T., Yagci H., Nomura S., Yamauchi A., Sasaki K., Deguchi K.,  
RA Shimizu Y., Bronson R.T., Gao Y.-H., Inada M., Sato M., Okamoto R.,  
RA Kitamura Y., Yoshiki S., Kishimoto T. ;  
RT "Targeted disruption of Chf1 results in a complete lack of bone  
RT formation owing to maturational arrest of osteoblasts." ;  
RL Cell 89:755-764(1997).  
RN [9]  
RP PHOSPHORYLATION.  
RX MEDLINE-20127938 ; PubMed-10660618 ;  
RA Xiao G., Jiang D., Thomas P., Benson M.D., Guan K., Karsenty G.,  
RA Franceschi R.T. ;  
RT "MAPK pathways activate and phosphorylate the osteoblast-specific  
RT transcription factor Cbfa1." ;  
RL J. Biol. Chem. 275:4453-4459(2000).  
RN [10]  
RP FUNCTION: Transcription factor involved in osteoblastic

[illegible]

```

FT FT VARSPLIC 428 607 ISOFORM 7).
FT FT TESSRSPNPMKHPATRTTTPYTSGLMSATTHHTL
FT FT PYPSSOSOSGPFQTSSTPYLYGTSSASYPMPVGGDR
FT FT SPSRWVPCTTSSNGSTLNPMLPNONDGVADGSHSSPT
FT FT VLNSGMRDESVWRY -> SEPTLDSSTTLFSSSEP
FT FT GPSTALPSSSSCEPQPSPPMLPPLLOPSTASTVAP
FT FT CVPRTGLTIVTSSPEAPHLVDMMSCPTATSPGVRKD
FT FT HRRPOTMAPAPALASERHSGHAGARDHAHEPETSXP
FT FT CAPPAATLEASVGDILVELRTMNGHDIKALTKLASS
FT FT LVPOSQPVEAPDAN (IN ISOFORM 8 AND ISOFORM
FT FT 9).
FT FT VARSPLIC 440 607 MISSING (IN ISOFORM 7).
FT FT CONFLICT 266 266 A -> S (IN REF. 4; AAC78626).
FT FT CONFLICT 280 280 G -> S (IN REF. 4; AAC78626).
FT FT CONFLICT 373 373 D -> N (IN REF. 4; AAC78626).
FT FT CONFLICT 375 375 M -> T (IN REF. 4; AAC78626).
FT FT CONFLICT 396 396 M -> L (IN REF. 4; AAC78626).
FT FT CONFLICT 459 459 P -> L (IN REF. 4; AAC78626).
FT FT CONFLICT 472 472 R -> P (IN REF. 4; AAC78626).
FT FT SEQUENCE 607 AA: 66204 MW: E87A4437ED19EE0E CRC64;

Query Match 74.4%; Score 32; DB 1; Length 607;
Best Local Similarity 62.5%; Pred. No. 83;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 POGQKVT 8
DB 365 POGQSQIT 372

RESULT 10
AAC2_HUMAN
ID AAC2_HUMAN STANDARD; PRT; 894 AA.
AC P35609;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Alpha-actinin 2 (Alpha actinin skeletal muscle isoform 2) (F-actin
DE cross linking protein).
GN ACTN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RC MEDLINE=92250531; PubMed=1339456;
RC Beggs A.H., Byers T.J., Knoll J.H.M., Boyce F.M., Bruns G.A.P.,
RC Kunzel L.W.;
RT "Cloning and characterization of two human skeletal muscle alpha-
RT actinin genes located on chromosomes 1 and 11."
RL J. Biol. Chem. 267:9281-9288(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20018005; PubMed=10548523;
RA Tiso N., Majetti M., Stanchi F., Rampazzo A., Zimbelli R., Nava A.,
RA Danelli G.A.;
RT "Fine mapping and genomic structure of ACTN2, the human gene coding
RT for the sarcomeric isoform of alpha-actinin-2, expressed in skeletal
RT and cardiac muscle."
RL Blochem. Biophys. Res. Commun. 265:256-259(1999).
RT ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING
RT PROTEIN.
CC -1- SUBUNIT: Homodimer, antiparallel. Also forms heterodimers with
CC ACTN3. Interacts with ADAM12.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BOTH SKELETAL AND CARDIAC MUSCLE.
CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.

```

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M66406; AAA51583.1; -
DR EMBL; M86804; AAA51584.1; -
DR EMBL; AJ249756; CAB61269.1; -
DR EMBL; AJ249757; CAB61269.1; JOINED.
DR EMBL; AJ249758; CAB61269.1; JOINED.
DR EMBL; AJ249759; CAB61269.1; JOINED.
DR EMBL; AJ249760; CAB61269.1; JOINED.
DR EMBL; AJ249761; CAB61269.1; JOINED.
DR EMBL; AJ249762; CAB61269.1; JOINED.
DR EMBL; AJ249763; CAB61269.1; JOINED.
DR EMBL; AJ249764; CAB61269.1; JOINED.
DR EMBL; AJ249765; CAB61269.1; JOINED.
DR EMBL; AJ249766; CAB61269.1; JOINED.
DR EMBL; AJ249767; CAB61269.1; JOINED.
DR EMBL; AJ249768; CAB61269.1; JOINED.
DR EMBL; AJ249769; CAB61269.1; JOINED.
DR EMBL; AJ249770; CAB61269.1; JOINED.
DR EMBL; AJ249771; CAB61269.1; JOINED.
DR EMBL; AJ249772; CAB61269.1; JOINED.
DR EMBL; AJ249773; CAB61269.1; JOINED.
DR EMBL; AJ249774; CAB61269.1; JOINED.
DR EMBL; AJ249775; CAB61269.1; JOINED.
DR EMBL; AJ249776; CAB61269.1; JOINED.
DR PIR; A40199; A40199.
DR HSSP; Q01082; IBKR.
DR MIM; 102573; -
DR InterPro: IPR001589; Actinin_act_bind.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR002017; Spectrin.
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00036; efhand; 2.
DR Pfam; PF00435; spectrin; 4.
DR SMART; SM00033; CH; 2.
DR SMART; SM00054; brp; 2.
DR SMART; SM00150; SPEC; 2.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS50021; CH; 2.
KW Actin-binding; Repeat; Multigene family.
FT DOMAIN 1 254 ACTIN-BINDING.
FT DOMAIN 2 254 ACTIN-BINDING.
FT DOMAIN 3 254 ACTIN-BINDING.
FT DOMAIN 4 254 ACTIN-BINDING.
FT DOMAIN 5 254 ACTIN-BINDING.
FT DOMAIN 6 254 ACTIN-BINDING.
FT DOMAIN 7 254 ACTIN-BINDING.
FT DOMAIN 8 254 ACTIN-BINDING.
FT DOMAIN 9 254 ACTIN-BINDING.
FT DOMAIN 10 254 ACTIN-BINDING.
FT DOMAIN 11 254 ACTIN-BINDING.
FT DOMAIN 12 254 ACTIN-BINDING.
FT DOMAIN 13 254 ACTIN-BINDING.
FT DOMAIN 14 254 ACTIN-BINDING.
FT DOMAIN 15 254 ACTIN-BINDING.
FT DOMAIN 16 254 ACTIN-BINDING.
FT DOMAIN 17 254 ACTIN-BINDING.
FT DOMAIN 18 254 ACTIN-BINDING.
FT DOMAIN 19 254 ACTIN-BINDING.
FT DOMAIN 20 254 ACTIN-BINDING.
FT DOMAIN 21 254 ACTIN-BINDING.
FT DOMAIN 22 254 ACTIN-BINDING.
FT DOMAIN 23 254 ACTIN-BINDING.
FT DOMAIN 24 254 ACTIN-BINDING.
FT DOMAIN 25 254 ACTIN-BINDING.
FT DOMAIN 26 254 ACTIN-BINDING.
FT DOMAIN 27 254 ACTIN-BINDING.
FT DOMAIN 28 254 ACTIN-BINDING.
FT DOMAIN 29 254 ACTIN-BINDING.
FT DOMAIN 30 254 ACTIN-BINDING.
FT DOMAIN 31 254 ACTIN-BINDING.
FT DOMAIN 32 254 ACTIN-BINDING.
FT DOMAIN 33 254 ACTIN-BINDING.
FT DOMAIN 34 254 ACTIN-BINDING.
FT DOMAIN 35 254 ACTIN-BINDING.
FT DOMAIN 36 254 ACTIN-BINDING.
FT DOMAIN 37 254 ACTIN-BINDING.
FT DOMAIN 38 254 ACTIN-BINDING.
FT DOMAIN 39 254 ACTIN-BINDING.
FT DOMAIN 40 254 ACTIN-BINDING.
FT DOMAIN 41 254 ACTIN-BINDING.
FT DOMAIN 42 254 ACTIN-BINDING.
FT DOMAIN 43 254 ACTIN-BINDING.
FT DOMAIN 44 254 ACTIN-BINDING.
FT DOMAIN 45 254 ACTIN-BINDING.
FT DOMAIN 46 254 ACTIN-BINDING.
FT DOMAIN 47 254 ACTIN-BINDING.
FT DOMAIN 48 254 ACTIN-BINDING.
FT DOMAIN 49 254 ACTIN-BINDING.
FT DOMAIN 50 254 ACTIN-BINDING.
FT DOMAIN 51 254 ACTIN-BINDING.
FT DOMAIN 52 254 ACTIN-BINDING.
FT DOMAIN 53 254 ACTIN-BINDING.
FT DOMAIN 54 254 ACTIN-BINDING.
FT DOMAIN 55 254 ACTIN-BINDING.
FT DOMAIN 56 254 ACTIN-BINDING.
FT DOMAIN 57 254 ACTIN-BINDING.
FT DOMAIN 58 254 ACTIN-BINDING.
FT DOMAIN 59 254 ACTIN-BINDING.
FT DOMAIN 60 254 ACTIN-BINDING.
FT DOMAIN 61 254 ACTIN-BINDING.
FT DOMAIN 62 254 ACTIN-BINDING.
FT DOMAIN 63 254 ACTIN-BINDING.
FT DOMAIN 64 254 ACTIN-BINDING.
FT DOMAIN 65 254 ACTIN-BINDING.
FT DOMAIN 66 254 ACTIN-BINDING.
FT DOMAIN 67 254 ACTIN-BINDING.
FT DOMAIN 68 254 ACTIN-BINDING.
FT DOMAIN 69 254 ACTIN-BINDING.
FT DOMAIN 70 254 ACTIN-BINDING.
FT DOMAIN 71 254 ACTIN-BINDING.
FT DOMAIN 72 254 ACTIN-BINDING.
FT DOMAIN 73 254 ACTIN-BINDING.
FT DOMAIN 74 254 ACTIN-BINDING.
FT DOMAIN 75 254 ACTIN-BINDING.
FT DOMAIN 76 254 ACTIN-BINDING.
FT DOMAIN 77 254 ACTIN-BINDING.
FT DOMAIN 78 254 ACTIN-BINDING.
FT DOMAIN 79 254 ACTIN-BINDING.
FT DOMAIN 80 254 ACTIN-BINDING.
FT DOMAIN 81 254 ACTIN-BINDING.
FT DOMAIN 82 254 ACTIN-BINDING.
FT DOMAIN 83 254 ACTIN-BINDING.
FT DOMAIN 84 254 ACTIN-BINDING.
FT DOMAIN 85 254 ACTIN-BINDING.
FT DOMAIN 86 254 ACTIN-BINDING.
FT DOMAIN 87 254 ACTIN-BINDING.
FT DOMAIN 88 254 ACTIN-BINDING.
FT DOMAIN 89 254 ACTIN-BINDING.
FT DOMAIN 90 254 ACTIN-BINDING.
FT DOMAIN 91 254 ACTIN-BINDING.
FT DOMAIN 92 254 ACTIN-BINDING.
FT DOMAIN 93 254 ACTIN-BINDING.
FT DOMAIN 94 254 ACTIN-BINDING.
FT DOMAIN 95 254 ACTIN-BINDING.
FT DOMAIN 96 254 ACTIN-BINDING.
FT DOMAIN 97 254 ACTIN-BINDING.
FT DOMAIN 98 254 ACTIN-BINDING.
FT DOMAIN 99 254 ACTIN-BINDING.
FT DOMAIN 100 254 ACTIN-BINDING.
FT DOMAIN 101 254 ACTIN-BINDING.
FT DOMAIN 102 254 ACTIN-BINDING.
FT DOMAIN 103 254 ACTIN-BINDING.
FT DOMAIN 104 254 ACTIN-BINDING.
FT DOMAIN 105 254 ACTIN-BINDING.
FT DOMAIN 106 254 ACTIN-BINDING.
FT DOMAIN 107 254 ACTIN-BINDING.
FT DOMAIN 108 254 ACTIN-BINDING.
FT DOMAIN 109 254 ACTIN-BINDING.
FT DOMAIN 110 254 ACTIN-BINDING.
FT DOMAIN 111 254 ACTIN-BINDING.
FT DOMAIN 112 254 ACTIN-BINDING.
FT DOMAIN 113 254 ACTIN-BINDING.
FT DOMAIN 114 254 ACTIN-BINDING.
FT DOMAIN 115 254 ACTIN-BINDING.
FT DOMAIN 116 254 ACTIN-BINDING.
FT DOMAIN 117 254 ACTIN-BINDING.
FT DOMAIN 118 254 ACTIN-BINDING.
FT DOMAIN 119 254 ACTIN-BINDING.
FT DOMAIN 120 254 ACTIN-BINDING.
FT DOMAIN 121 254 ACTIN-BINDING.
FT DOMAIN 122 254 ACTIN-BINDING.
FT DOMAIN 123 254 ACTIN-BINDING.
FT DOMAIN 124 254 ACTIN-BINDING.
FT DOMAIN 125 254 ACTIN-BINDING.
FT DOMAIN 126 254 ACTIN-BINDING.
FT DOMAIN 127 254 ACTIN-BINDING.
FT DOMAIN 128 254 ACTIN-BINDING.
FT DOMAIN 129 254 ACTIN-BINDING.
FT DOMAIN 130 254 ACTIN-BINDING.
FT DOMAIN 131 254 ACTIN-BINDING.
FT DOMAIN 132 254 ACTIN-BINDING.
FT DOMAIN 133 254 ACTIN-BINDING.
FT DOMAIN 134 254 ACTIN-BINDING.
FT DOMAIN 135 254 ACTIN-BINDING.
FT DOMAIN 136 254 ACTIN-BINDING.
FT DOMAIN 137 254 ACTIN-BINDING.
FT DOMAIN 138 254 ACTIN-BINDING.
FT DOMAIN 139 254 ACTIN-BINDING.
FT DOMAIN 140 254 ACTIN-BINDING.
FT DOMAIN 141 254 ACTIN-BINDING.
FT DOMAIN 142 254 ACTIN-BINDING.
FT DOMAIN 143 254 ACTIN-BINDING.
FT DOMAIN 144 254 ACTIN-BINDING.
FT DOMAIN 145 254 ACTIN-BINDING.
FT DOMAIN 146 254 ACTIN-BINDING.
FT DOMAIN 147 254 ACTIN-BINDING.
FT DOMAIN 148 254 ACTIN-BINDING.
FT DOMAIN 149 254 ACTIN-BINDING.
FT DOMAIN 150 254 ACTIN-BINDING.
FT DOMAIN 151 254 ACTIN-BINDING.
FT DOMAIN 152 254 ACTIN-BINDING.
FT DOMAIN 153 254 ACTIN-BINDING.
FT DOMAIN 154 254 ACTIN-BINDING.
FT DOMAIN 155 254 ACTIN-BINDING.
FT DOMAIN 156 254 ACTIN-BINDING.
FT DOMAIN 157 254 ACTIN-BINDING.
FT DOMAIN 158 254 ACTIN-BINDING.
FT DOMAIN 159 254 ACTIN-BINDING.
FT DOMAIN 160 254 ACTIN-BINDING.
FT DOMAIN 161 254 ACTIN-BINDING.
FT DOMAIN 162 254 ACTIN-BINDING.
FT DOMAIN 163 254 ACTIN-BINDING.
FT DOMAIN 164 254 ACTIN-BINDING.
FT DOMAIN 165 254 ACTIN-BINDING.
FT DOMAIN 166 254 ACTIN-BINDING.
FT DOMAIN 167 254 ACTIN-BINDING.
FT DOMAIN 168 254 ACTIN-BINDING.
FT DOMAIN 169 254 ACTIN-BINDING.
FT DOMAIN 170 254 ACTIN-BINDING.
FT DOMAIN 171 254 ACTIN-BINDING.
FT DOMAIN 172 254 ACTIN-BINDING.
FT DOMAIN 173 254 ACTIN-BINDING.
FT DOMAIN 174 254 ACTIN-BINDING.
FT DOMAIN 175 254 ACTIN-BINDING.
FT DOMAIN 176 254 ACTIN-BINDING.
FT DOMAIN 177 254 ACTIN-BINDING.
FT DOMAIN 178 254 ACTIN-BINDING.
FT DOMAIN 179 254 ACTIN-BINDING.
FT DOMAIN 180 254 ACTIN-BINDING.
FT DOMAIN 181 254 ACTIN-BINDING.
FT DOMAIN 182 254 ACTIN-BINDING.
FT DOMAIN 183 254 ACTIN-BINDING.
FT DOMAIN 184 254 ACTIN-BINDING.
FT DOMAIN 185 254 ACTIN-BINDING.
FT DOMAIN 186 254 ACTIN-BINDING.
FT DOMAIN 187 254 ACTIN-BINDING.
FT DOMAIN 188 254 ACTIN-BINDING.
FT DOMAIN 189 254 ACTIN-BINDING.
FT DOMAIN 190 254 ACTIN-BINDING.
FT DOMAIN 191 254 ACTIN-BINDING.
FT DOMAIN 192 254 ACTIN-BINDING.
FT DOMAIN 193 254 ACTIN-BINDING.
FT DOMAIN 194 254 ACTIN-BINDING.
FT DOMAIN 195 254 ACTIN-BINDING.
FT DOMAIN 196 254 ACTIN-BINDING.
FT DOMAIN 197 254 ACTIN-BINDING.
FT DOMAIN 198 254 ACTIN-BINDING.
FT DOMAIN 199 254 ACTIN-BINDING.
FT DOMAIN 200 254 ACTIN-BINDING.
FT DOMAIN 201 254 ACTIN-BINDING.
FT DOMAIN 202 254 ACTIN-BINDING.
FT DOMAIN 203 254 ACTIN-BINDING.
FT DOMAIN 204 254 ACTIN-BINDING.
FT DOMAIN 205 254 ACTIN-BINDING.
FT DOMAIN 206 254 ACTIN-BINDING.
FT DOMAIN 207 254 ACTIN-BINDING.
FT DOMAIN 208 254 ACTIN-BINDING.
FT DOMAIN 209 254 ACTIN-BINDING.
FT DOMAIN 210 254 ACTIN-BINDING.
FT DOMAIN 211 254 ACTIN-BINDING.
FT DOMAIN 212 254 ACTIN-BINDING.
FT DOMAIN 213 254 ACTIN-BINDING.
FT DOMAIN 214 254 ACTIN-BINDING.
FT DOMAIN 215 254 ACTIN-BINDING.
FT DOMAIN 216 254 ACTIN-BINDING.
FT DOMAIN 217 254 ACTIN-BINDING.
FT DOMAIN 218 254 ACTIN-BINDING.
FT DOMAIN 219 254 ACTIN-BINDING.
FT DOMAIN 220 254 ACTIN-BINDING.
FT DOMAIN 221 254 ACTIN-BINDING.
FT DOMAIN 222 254 ACTIN-BINDING.
FT DOMAIN 223 254 ACTIN-BINDING.
FT DOMAIN 224 254 ACTIN-BINDING.
FT DOMAIN 225 254 ACTIN-BINDING.
FT DOMAIN 226 254 ACTIN-BINDING.
FT DOMAIN 227 254 ACTIN-BINDING.
FT DOMAIN 228 254 ACTIN-BINDING.
FT DOMAIN 229 254 ACTIN-BINDING.
FT DOMAIN 230 254 ACTIN-BINDING.
FT DOMAIN 231 254 ACTIN-BINDING.
FT DOMAIN 232 254 ACTIN-BINDING.
FT DOMAIN 233 254 ACTIN-BINDING.
FT DOMAIN 234 254 ACTIN-BINDING.
FT DOMAIN 235 254 ACTIN-BINDING.
FT DOMAIN 236 254 ACTIN-BINDING.
FT DOMAIN 237 254 ACTIN-BINDING.
FT DOMAIN 238 254 ACTIN-BINDING.
FT DOMAIN 239 254 ACTIN-BINDING.
FT DOMAIN 240 254 ACTIN-BINDING.
FT DOMAIN 241 254 ACTIN-BINDING.
FT DOMAIN 242 254 ACTIN-BINDING.
FT DOMAIN 243 254 ACTIN-BINDING.
FT DOMAIN 244 254 ACTIN-BINDING.
FT DOMAIN 245 254 ACTIN-BINDING.
FT DOMAIN 246 254 ACTIN-BINDING.
FT DOMAIN 247 254 ACTIN-BINDING.
FT DOMAIN 248 254 ACTIN-BINDING.
FT DOMAIN 249 254 ACTIN-BINDING.
FT DOMAIN 250 254 ACTIN-BINDING.
FT DOMAIN 251 254 ACTIN-BINDING.
FT DOMAIN 252 254 ACTIN-BINDING.
FT DOMAIN 253 254 ACTIN-BINDING.
FT DOMAIN 254 254 ACTIN-BINDING.
FT DOMAIN 255 254 ACTIN-BINDING.
FT DOMAIN 256 254 ACTIN-BINDING.
FT DOMAIN 257 254 ACTIN-BINDING.
FT DOMAIN 258 254 ACTIN-BINDING.
FT DOMAIN 259 254 ACTIN-BINDING.
FT DOMAIN 260 254 ACTIN-BINDING.
FT DOMAIN 261 254 ACTIN-BINDING.
FT DOMAIN 262 254 ACTIN-BINDING.
FT DOMAIN 263 254 ACTIN-BINDING.
FT DOMAIN 264 254 ACTIN-BINDING.
FT DOMAIN 265 254 ACTIN-BINDING.
FT DOMAIN 266 254 ACTIN-BINDING.
FT DOMAIN 267 254 ACTIN-BINDING.
FT DOMAIN 268 254 ACTIN-BINDING.
FT DOMAIN 269 254 ACTIN-BINDING.
FT DOMAIN 270 254 ACTIN-BINDING.
FT DOMAIN 271 254 ACTIN-BINDING.
FT DOMAIN 272 254 ACTIN-BINDING.
FT DOMAIN 273 254 ACTIN-BINDING.
FT DOMAIN 274 254 ACTIN-BINDING.
FT DOMAIN 275 254 ACTIN-BINDING.
FT DOMAIN 276 254 ACTIN-BINDING.
FT DOMAIN 277 254 ACTIN-BINDING.
FT DOMAIN 278 254 ACTIN-BINDING.
FT DOMAIN 279 254 ACTIN-BINDING.
FT DOMAIN 280 254 ACTIN-BINDING.
FT DOMAIN 281 254 ACTIN-BINDING.
FT DOMAIN 282 254 ACTIN-BINDING.
FT DOMAIN 283 254 ACTIN-BINDING.
FT DOMAIN 284 254 ACTIN-BINDING.
FT DOMAIN 285 254 ACTIN-BINDING.
FT DOMAIN 286 254 ACTIN-BINDING.
FT DOMAIN 287 254 ACTIN-BINDING.
FT DOMAIN 288 254 ACTIN-BINDING.
FT DOMAIN 289 254 ACTIN-BINDING.
FT DOMAIN 290 254 ACTIN-BINDING.
FT DOMAIN 291 254 ACTIN-BINDING.
FT DOMAIN 292 254 ACTIN-BINDING.
FT DOMAIN 293 254 ACTIN-BINDING.
FT DOMAIN 294 254 ACTIN-BINDING.
FT DOMAIN 295 254 ACTIN-BINDING.
FT DOMAIN 296 254 ACTIN-BINDING.
FT DOMAIN 297 254 ACTIN-BINDING.
FT DOMAIN 298 254 ACTIN-BINDING.
FT DOMAIN 299 254 ACTIN-BINDING.
FT DOMAIN 300 254 ACTIN-BINDING.
FT DOMAIN 301 254 ACTIN-BINDING.
FT DOMAIN 302 254 ACTIN-BINDING.
FT DOMAIN 303 254 ACTIN-BINDING.
FT DOMAIN 304 254 ACTIN-BINDING.
FT DOMAIN 305 254 ACTIN-BINDING.
FT DOMAIN 306 254 ACTIN-BINDING.
FT DOMAIN 307 254 ACTIN-BINDING.
FT DOMAIN 308 254 ACTIN-BINDING.
FT DOMAIN 309 254 ACTIN-BINDING.
FT DOMAIN 310 254 ACTIN-BINDING.
FT DOMAIN 311 254 ACTIN-BINDING.
FT DOMAIN 312 254 ACTIN-BINDING.
FT DOMAIN 313 254 ACTIN-BINDING.
FT DOMAIN 314 254 ACTIN-BINDING.
FT DOMAIN 315 254 ACTIN-BINDING.
FT DOMAIN 316 254 ACTIN-BINDING.
FT DOMAIN 317 254 ACTIN-BINDING.
FT DOMAIN 318 254 ACTIN-BINDING.
FT DOMAIN 319 254 ACTIN-BINDING.
FT DOMAIN 320 254 ACTIN-BINDING.
FT DOMAIN 321 254 ACTIN-BINDING.
FT DOMAIN 322 254 ACTIN-BINDING.
FT DOMAIN 323 254 ACTIN-BINDING.
FT DOMAIN 324 254 ACTIN-BINDING.
FT DOMAIN 325 254 ACTIN-BINDING.
FT DOMAIN 326 254 ACTIN-BINDING.
FT DOMAIN 327 254 ACTIN-BINDING.
FT DOMAIN 328 254 ACTIN-BINDING.
FT DOMAIN 329 254 ACTIN-BINDING.
FT DOMAIN 330 254 ACTIN-BINDING.
FT DOMAIN 331 254 ACTIN-BINDING.
FT DOMAIN 332 254 ACTIN-BINDING.
FT DOMAIN 333 254 ACTIN-BINDING.
FT DOMAIN 334 254 ACTIN-BINDING.
FT DOMAIN 335 254 ACTIN-BINDING.
FT DOMAIN 336 254 ACTIN-BINDING.
FT DOMAIN 337 254 ACTIN-BINDING.
FT DOMAIN 338 254 ACTIN-BINDING.
FT DOMAIN 339 254 ACTIN-BINDING.
FT DOMAIN 340 254 ACTIN-BINDING.
FT DOMAIN 341 254 ACTIN-BINDING.
FT DOMAIN 342 254 ACTIN-BINDING.
FT DOMAIN 343 254 ACTIN-BINDING.
FT DOMAIN 344 254 ACTIN-BINDING.
FT DOMAIN 345 254 ACTIN-BINDING.
FT DOMAIN 346 254 ACTIN-BINDING.
FT DOMAIN 347 254 ACTIN-BINDING.
FT DOMAIN 348 254 ACTIN-BINDING.
FT DOMAIN 349 254 ACTIN-BINDING.
FT DOMAIN 350 254 ACTIN-BINDING.
FT DOMAIN 351 254 ACTIN-BINDING.
FT DOMAIN 352 254 ACTIN-BINDING.
FT DOMAIN 353 254 ACTIN-BINDING.
FT DOMAIN 354 254 ACTIN-BINDING.
FT DOMAIN 355 254 ACTIN-BINDING.
FT DOMAIN 356 254 ACTIN-BINDING.
FT DOMAIN 357 254 ACTIN-BINDING.
FT DOMAIN 358 254 ACTIN-BINDING.
FT DOMAIN 359 254 ACTIN-BINDING.
FT DOMAIN 360 254 ACTIN-BINDING.
FT DOMAIN 361 254 ACTIN-BINDING.
FT DOMAIN 362 254 ACTIN-BINDING.
FT DOMAIN 363 254 ACTIN-BINDING.
FT DOMAIN 364 254 ACTIN-BINDING.
FT DOMAIN 365 254 ACTIN-BINDING.
FT DOMAIN 366 254 ACTIN-BINDING.
FT DOMAIN 367 254 ACTIN-BINDING.
FT DOMAIN 368 254 ACTIN-BINDING.
FT DOMAIN 369 254 ACTIN-BINDING.
FT DOMAIN 370 254 ACTIN-BINDING.
FT DOMAIN 371 254 ACTIN-BINDING.
FT DOMAIN 372 254 ACTIN-BINDING.
FT DOMAIN 373 254 ACTIN-BINDING.
FT DOMAIN 374 254 ACTIN-BINDING.
FT DOMAIN 375 254 ACTIN-BINDING.
FT DOMAIN 376 254 ACTIN-BINDING.
FT DOMAIN 377 254 ACTIN-BINDING.
FT DOMAIN 378 254 ACTIN-BINDING.
FT DOMAIN 379 254 ACTIN-BINDING.
FT DOMAIN 380 254 ACTIN-BINDING.
FT DOMAIN 381 254 ACTIN-BINDING.
FT DOMAIN 382 254 ACTIN-BINDING.
FT DOMAIN 383 254 ACTIN-BINDING.
FT DOMAIN 384 254 ACTIN-BINDING.
FT DOMAIN 385 254 ACTIN-BINDING.
FT DOMAIN 386 254 ACTIN-BINDING.
FT DOMAIN 387 254 ACTIN-BINDING.
FT DOMAIN 388 254 ACTIN-BINDING.
FT DOMAIN 389 254 ACTIN-BINDING.
FT DOMAIN 390 254 ACTIN-BINDING.
FT DOMAIN 391 254 ACTIN-BINDING.
FT DOMAIN 392 254 ACTIN-BINDING.
FT DOMAIN 393 254 ACTIN-BINDING.
FT DOMAIN 394 254 ACTIN-BINDING.
FT DOMAIN 395 254 ACTIN-BINDING.
FT DOMAIN 396 254 ACTIN-BINDING.
FT DOMAIN 397 254 ACTIN-BINDING.
FT DOMAIN 398 254 ACTIN-BINDING.
FT DOMAIN 399 254 ACTIN-BINDING.
FT DOMAIN 400 254 ACTIN-BINDING.
FT DOMAIN 401 254 ACTIN-BINDING.
FT DOMAIN 402 254 ACTIN-BINDING.
FT DOMAIN 403 254 ACTIN-BINDING.
FT DOMAIN 404 254 ACTIN-BINDING.
FT DOMAIN 405 254 ACTIN-BINDING.
FT DOMAIN 406 254 ACTIN-BINDING.
FT DOMAIN 407 254 ACTIN-BINDING.
FT DOMAIN 408 254 ACTIN-BINDING.
FT DOMAIN 409 254 ACTIN-BINDING.
FT DOMAIN 410 254 ACTIN-BINDING.
FT DOMAIN 411 254 ACTIN-BINDING.
FT DOMAIN 412 254 ACTIN-BINDING.
FT DOMAIN 413 254 ACTIN-BINDING.
FT DOMAIN 414 254 ACTIN-BINDING.
FT DOMAIN 415 254 ACTIN-BINDING.
FT DOMAIN 416 254 ACTIN-BINDING.
FT DOMAIN 417 254 ACTIN-BINDING.
FT DOMAIN 418 254 ACTIN-BINDING.
FT DOMAIN 419 254 ACTIN-BINDING.
FT DOMAIN 420 254 ACTIN-BINDING.
FT DOMAIN 421 254 ACTIN-BINDING.
FT DOMAIN 422 254 ACTIN-BINDING.
FT DOMAIN 423 254 ACTIN-BINDING.
FT DOMAIN 424 254 ACTIN-BINDING.
FT DOMAIN 425 254 ACTIN-BINDING.
FT DOMAIN 426 254 ACTIN-BINDING.
FT DOMAIN 427 254 ACTIN-BINDING.
FT DOMAIN 428 254 ACTIN-BINDING.
FT DOMAIN 429 254 ACTIN-BINDING.
FT DOMAIN 430 254 ACTIN-BINDING.
FT DOMAIN 431 254 ACTIN-BINDING.
FT DOMAIN 432 254 ACTIN-BINDING.
FT DOMAIN 433 254 ACTIN-BINDING.
FT DOMAIN 434 254 ACTIN-BINDING.
FT DOMAIN 435 254 ACTIN-BINDING.
FT DOMAIN 436 254 ACTIN-BINDING.
FT DOMAIN 437 254 ACTIN-BINDING.
FT DOMAIN 438 254 ACTIN-BINDING.
FT DOMAIN 439 254 ACTIN-BINDING.
FT DOMAIN 440 254 ACTIN-BINDING.
FT DOMAIN 441 254 ACTIN-BINDING.
FT DOMAIN 442 254 ACTIN-BINDING.
FT DOMAIN 443 254 ACTIN-BINDING.
FT DOMAIN 444 254 ACTIN-BINDING.
FT DOMAIN 445 254 ACTIN-BINDING.
FT DOMAIN 446 254 ACTIN-BINDING.
FT DOMAIN 447 254 ACTIN-BINDING.
FT DOMAIN 448 254 ACTIN-BINDING.
FT DOMAIN 449 254 ACTIN-BINDING.
FT DOMAIN 450 254 ACTIN-BINDING.
FT DOMAIN 451 254 ACTIN-BINDING.
FT DOMAIN 452 254 ACTIN-BINDING.
FT DOMAIN 453 254 ACTIN-BINDING.
FT DOMAIN 454 254 ACTIN-BINDING.
FT DOMAIN 455 254 ACTIN-BINDING.
FT DOMAIN 456 254 ACTIN-BINDING.
FT DOMAIN 457 254 ACTIN-BINDING.
FT DOMAIN 458 254 ACTIN-BINDING.
FT DOMAIN 459 254 ACTIN-BINDING.
FT DOMAIN 460 254 ACTIN-BINDING.
FT DOMAIN 461 254 ACTIN-BINDING.
FT DOMAIN 462 254 ACTIN-BINDING.
FT DOMAIN 463 254 ACTIN-BINDING.
FT DOMAIN 464 254 ACTIN-BINDING.
FT DOMAIN 465 254 ACTIN-BINDING.
FT DOMAIN 466 254 ACTIN-BINDING.
FT DOMAIN 467 254 ACTIN-BINDING.
FT DOMAIN 468 254 ACTIN-BINDING.
FT DOMAIN 469 254 ACTIN-BINDING.
FT DOMAIN 470 254 ACTIN-BINDING.
FT DOMAIN 471 254 ACTIN-BINDING.
FT DOMAIN 472 254 ACTIN-BINDING.
FT DOMAIN 473 254 ACTIN-BINDING.
FT DOMAIN 474 254 ACTIN-BINDING.
FT DOMAIN 475 254 ACTIN-BINDING.
FT DOMAIN 476 254 ACTIN-BINDING.
FT DOMAIN 477 254 ACTIN-BINDING.
FT DOMAIN 478 254 ACTIN-BINDING.
FT DOMAIN 479 254 ACTIN-BINDING.
FT DOMAIN 480 254 ACTIN-BINDING.
FT DOMAIN 481 254 ACTIN-BINDING.
FT DOMAIN 482 254 ACTIN-BINDING.
FT DOMAIN 483 254 ACTIN-BINDING.
FT DOMAIN 484 254 ACTIN-BINDING.
FT DOMAIN 485 254 ACTIN-BINDING.
FT DOMAIN 486 254 ACTIN-BINDING.
FT DOMAIN 487 254 ACTIN-BINDING.
FT DOMAIN 488 254 ACTIN-BINDING.
FT DOMAIN 489 254 ACTIN-BINDING.
FT DOMAIN 490 254 ACTIN-BINDING.
FT DOMAIN 491 254 ACTIN-BINDING.
FT DOMAIN 492 254 ACTIN-BINDING.
FT DOMAIN 493 254 ACTIN-BINDING.
FT DOMAIN 494 254 ACTIN-BINDING.
FT DOMAIN 495 254 ACTIN-BINDING.
FT DOMAIN 496 254 ACTIN-BINDING.
FT DOMAIN 497 254 ACTIN-BINDING.
FT DOMAIN 498 254 ACTIN-BINDING.
FT DOMAIN 499 254 ACTIN-BINDING.
FT DOMAIN 500 254 ACTIN-BINDING.
FT DOMAIN 501 254 ACTIN-BINDING.
FT DOMAIN 502 254 ACTIN-BINDING.
FT DOMAIN 503 254 ACTIN-BINDING.
FT DOMAIN 504 254 ACTIN-BINDING.
FT DOMAIN 505 254 ACTIN-BINDING.
FT DOMAIN 506 254 ACTIN-BINDING.
FT DOMAIN 507 254 ACTIN-BINDING.
FT DOMAIN 508 254 ACTIN-BINDING.
FT DOMAIN 509 254 ACTIN-BINDING.
FT DOMAIN 510 254 ACTIN-BINDING.
FT DOMAIN 511 254 ACTIN-BINDING.
FT DOMAIN 512 254 ACTIN-BINDING.
FT DOMAIN 513 254 ACTIN-BINDING.
FT DOMAIN 514 254 ACTIN-BINDING.
FT DOMAIN 515 254 ACTIN-BINDING.
FT DOMAIN 516 254 ACTIN-BINDING.
FT DOMAIN 517 254 ACTIN-BINDING.
FT DOMAIN 518 254 ACTIN-BINDING.
FT DOMAIN 519 254 ACTIN-BINDING.
FT DOMAIN 520 254 ACTIN-BINDING.
FT DOMAIN 521 254 ACTIN-BINDING.
FT DOMAIN 522 254 ACTIN-BINDING.
FT DOMAIN 523 254 ACTIN-BINDING.
FT DOMAIN 524 254 ACTIN-BINDING.
FT DOMAIN 525 254 ACTIN-BINDING.
FT DOMAIN 526 254 ACTIN-BINDING.
FT DOMAIN 527 254 ACTIN-BINDING.
FT DOMAIN 528 254 ACTIN-BINDING.
FT DOMAIN 529 254 ACTIN-BINDING.
FT DOMAIN 530 254 ACTIN-BINDING.
FT DOMAIN 531 254 ACTIN-BINDING.
FT DOMAIN 532 254 ACTIN-BINDING.
FT DOMAIN 533 254 ACTIN-BINDING.
FT DOMAIN 534 254 ACTIN-BINDING.
FT DOMAIN 535 254 ACTIN-BINDING.
FT DOMAIN 536 254 ACTIN-BINDING.
FT DOMAIN 537 254 ACTIN-BINDING.
FT DOMAIN 538 254 ACTIN-BINDING.
FT DOMAIN 539 254 ACTIN-BINDING.
FT DOMAIN 540 254 ACTIN-BINDING.
FT DOMAIN 541 254 ACTIN-BINDING.
FT DOMAIN 542 254 ACTIN-BINDING.
FT DOMAIN 543 254 ACTIN-BINDING.
FT DOMAIN 544 254 ACTIN-BINDING.
FT DOMAIN 545 254 ACTIN-BINDING.
FT DOMAIN 546 254 ACTIN-BINDING.
FT DOMAIN 547 254 ACTIN-BINDING.
FT DOMAIN 548 254 ACTIN-BINDING.
FT DOMAIN 549 254 ACTIN-BINDING.
FT DOMAIN 550 254 ACTIN-BINDING.
FT DOMAIN 551 254 ACTIN-BINDING.
FT DOMAIN 552 254 ACTIN-BINDING.
FT DOMAIN 553 254 ACTIN-BINDING.
FT DOMAIN 554 254 ACTIN-BINDING.
FT DOMAIN 555 254 ACTIN-BINDING.
FT DOMAIN 556 254 ACTIN-BINDING.
FT DOMAIN 557 254 ACTIN-BINDING.
FT DOMAIN 558 254 ACTIN-BINDING.
FT DOMAIN 559 254 ACTIN-BINDING.
FT DOMAIN 560 254 ACTIN-BINDING.
FT DOMAIN 561 254 ACTIN-BINDING.
FT DOMAIN 562 254 ACTIN-BINDING.
FT DOMAIN 563 254 ACTIN-BINDING.
FT DOMAIN 564 254 ACTIN-BINDING.
FT DOMAIN 565 254 ACTIN-BINDING.
FT DOMAIN 566 254 ACTIN-BINDING.
FT DOMAIN 567 254 ACTIN-BINDING.
FT DOMAIN 568 254 ACTIN-BINDING.
FT DOMAIN 569 254 ACTIN-BINDING.
FT DOMAIN 570 254 ACTIN-BINDING.
FT DOMAIN 571 254 ACTIN-BINDING.
FT DOMAIN 572 254 ACTIN-BINDING.
FT DOMAIN 573 254 ACTIN-BINDING.
FT DOMAIN 574 254 ACTIN-BINDING.
FT DOMAIN 575 254 ACTIN-BINDING.
FT DOMAIN 576 254 ACTIN-BINDING.
FT DOMAIN 577 254 ACTIN-BINDING.
FT DOMAIN 578 254 ACTIN-BINDING.
FT DOMAIN 579 254 ACTIN-BINDING.
FT DOMAIN 580 254 ACTIN-BINDING.
FT DOMAIN 581 254 ACTIN-BINDING.
FT DOMAIN 582 254 ACTIN-BINDING.
FT DOMAIN 583 254 ACTIN-BINDING.
FT DOMAIN 584 254 ACTIN-BINDING.
FT DOMAIN 585 254 ACTIN-BINDING.
FT DOMAIN 586 254 ACTIN-BINDING
```

DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Alpha-actinin 2 (Alpha actinin skeletal muscle isoform 2) (F-actin cross linking protein).  
 GN ACTN2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skeletal muscle;  
 RX PubMed=11440986;  
 RA Mills M., Yang N., Weinberger R., Vander Woude D.L., Beggs A.H., Eastaugh S., North K.N.;  
 RT "Differential expression of the actin-binding proteins, alpha-actinin-2 and -3, in different species: implications for the evolution of functional redundancy.";  
 RL Hum. Mol. Genet. 10:1335-1346(2001).  
 [2]  
 RP INTERACTION WITH ADAM-12.  
 RX PubMed=10788519;  
 RA Galliano M.-F., Huet C., Flygellius J., Polgren A., Wewer U.M., Engvall E.;  
 RT "Binding of ADAM12, a marker of skeletal muscle regeneration, to the muscle-specific actin-binding protein, alpha-actinin-2, is required for myoblast fusion.";  
 RL J. Biol. Chem. 275:13933-13939(2000).  
 CC -1- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING PROTEIN (BY SIMILARITY).  
 CC -1- SUBUNIT: Homodimer, antiparallel. Also forms heterodimers with ACTN3. Interacts with ADAM12.  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOG (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: AF248643; AAF76325.1; -  
 CC MGD: MGI:109192; Actn2.  
 DR InterPro: IPR001589; Actinin\_act\_bind.  
 DR InterPro: IPR001715; Calponin\_hom.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00036; ehand; 2.  
 DR Pfam: PF00435; spectrin; 4.  
 DR SMART: SM00054; EFh; 2.  
 DR SMART: SM00053; CH; 2.  
 DR SMART: SM00150; SPEC; 2.  
 DR PROSITE: PS00019; ACTININ\_1; 1.  
 DR PROSITE: PS00020; ACTININ\_2; 1.  
 DR PROSITE: PS50021; CH; 2.  
 DR Actin-binding; Repeat; Multigene family.  
 KW Actin-binding; Repeat; Multigene family.  
 FT DOMAIN 1 254 ACTIN-BINDING.  
 FT DOMAIN 38 142 CH 1.  
 FT DOMAIN 151 254 CH 2.  
 FT REPEAT 281 391 SPECTRIN 1.  
 FT REPEAT 401 506 SPECTRIN 2.  
 FT REPEAT 516 627 SPECTRIN 3.  
 FT REPEAT 637 740 SPECTRIN 4.  
 FT CA\_BIND 766 777 EF-HAND 1 (POSSIBLY ANCESTRAL).  
 FT CA\_BIND 802 813 EF-HAND 2 (PROBABLY ANCESTRAL).  
 SEQUENCE 894 AA: 103653 MW: 4D707F59FB31C5AC CRC64;

Query Match 74.4%; Score 32; DB 1; Length 894;  
 Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 PGGGKRV 8  
 DB 803 PGGGKRV 810  
 RESULT 12  
 AAC2\_CHICK  
 ID AAC2\_CHICK STANDARD: PRT; 897 AA.  
 AC P20111;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Alpha-actinin 2 (Alpha actinin skeletal muscle isoform 2) (F-actin cross linking protein).  
 GN ACTN2.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
 NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89064821; PubMed=3197725;  
 RA Arimura C., Suzuki T., Yanagisawa M., Imamura M., Hamada Y., Masaki T.;  
 RT "Primary structure of chicken skeletal muscle and fibroblast alpha-actinins deduced from cDNA sequences.";  
 RL Eur. J. Biochem. 177:649-655(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92070385; PubMed=1720388;  
 RA Tokune Y., Goto S., Imamura M., Obinata T., Masaki T., Endo T.;  
 RT "Transfection of chicken skeletal muscle alpha-actinin cDNA into nonmuscle and myogenic cells: dimerization is not essential for alpha-actinin to bind to microfilaments.";  
 RL Exp. Cell Res. 197:158-167(1991).  
 CC -1- FUNCTION: F-ACTIN CROSS-LINKING PROTEIN WHICH IS THOUGHT TO ANCHOR ACTIN TO A VARIETY OF INTRACELLULAR STRUCTURES. THIS IS A BUNDLING PROTEIN.  
 CC -1- SUBUNIT: HOMODIMER, ANTIPARALLEL.  
 CC -1- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOG (CH) DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: X13874; CAA32078.1; -  
 CC EMBL: X59247; CAA41935.1; -  
 DR PIR: S02032; S02032.  
 DR PIR: S15481; S15481.  
 DR HSP: Q01082; 1BKR.  
 DR InterPro: IPR001589; Actinin\_act\_bind.  
 DR InterPro: IPR001715; Calponin\_hom.  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR002017; Spectrin.  
 DR Pfam: PF00307; CH; 2.  
 DR Pfam: PF00036; ehand; 2.  
 DR Pfam: PF00435; spectrin; 4.  
 DR SMART: SM00054; EFh; 2.  
 DR SMART: SM00053; CH; 2.  
 DR SMART: SM00150; SPEC; 2.

DR PROSITE; PS00019; ACTININ\_1; 1.  
DR PROSITE; PS00020; ACTININ\_2; 1.  
DR PROSITE; PS00021; CH\_2.  
DR PROSITE; PS00018; EF\_HAND; FALSE\_NEG.  
KW Actin-binding; Calcium-binding; Repeat; Multigene family.  
FT DOMAIN 1 257 ACTIN-BINDING.  
FT DOMAIN 1 145 CH 1.  
FT DOMAIN 154 257 CH 2.  
FT REPEAT 284 394 SPECTRIN 1.  
FT REPEAT 404 509 SPECTRIN 2.  
FT REPEAT 519 630 SPECTRIN 3.  
FT REPEAT 640 743 SPECTRIN 4.  
FT CA\_BIND 769 780 EF\_HAND 1 (POTENTIAL).  
FT CA\_BIND 805 816 EF\_HAND 2 (POTENTIAL).  
SQ SEQUENCE 897 AA; 104275 MW; F4FAC12F7F4C8834 CRC64;  
Query Match 74.4%; Score 32; DB 1; Length 897;  
Best Local Similarity 75.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
1 POGGKVT 8  
111111  
806 PGGGKVT 813  
RESULT 13  
FMA\_SERMA  
ID FMA\_SERMA STANDARD; PRT; 174 AA.  
AC P13421;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 01-JAN-1990 (Rel. 13, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE Fimbria A protein precursor.  
GN SMRA.  
OS Serratia marcescens.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC Serratia.  
OX NCBI\_TaxID=615;  
RN [1]  
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-37.  
RC STRAIN=US46;  
RX MEDLINE=88298666; PubMed=2900238;  
RA Mizuno Y., Nakabeppu Y., Sekiguchi M., Kawabata S.-I., Moriya T.,  
Amako K.;  
RT "Cloning and sequence of the gene encoding the major structural  
component of mannose-resistant fimbriae of Serratia marcescens.";  
RL J. Bacteriol. 170:3567-3574(1988).  
CC -1- FUNCTION: MAJOR STRUCTURAL COMPONENT OF MANNOSE-RESISTANT FIMBRIAE  
OF SERRATIA MARCESCENS.  
CC -1- SIMILARITY: BELONGS TO THE FIMA/PAPA FAMILY OF FIMBRIA PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
CC  
CC EMBL; M21161; AAA26576.1; -.  
DR PIR; A31096; A31096.  
DR InterPro; IPR000259; Fimbrin.  
DR Pfam; PF00419; Fimbrin; 1.  
KW Fimbrin; Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 174 FIMBRIA A PROTEIN.  
FT DISULFID 41 80 PROBABLE.  
SQ SEQUENCE 174 AA; 17771 MW; 0348B1003E113CA4 CRC64;  
Query Match 72.1%; Score 31; DB 1; Length 174;  
Best Local Similarity 85.7%; Pred. No. 40;  
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 OGCGKVT 8  
111111  
Db 25 OGCGKVT 31  
RESULT 14  
FLA5\_PYRO  
ID FLA5\_PYRO STANDARD; PRT; 255 AA.  
AC 058286;  
DT 30-MAY-2000 (Rel. 39, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Flagellin B5 precursor.  
GN FLA5 OR PH0551.  
OS Pyrococcus horikoshii.  
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.  
OX NCBI\_TaxID=53953;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OT3;  
RX MEDLINE=98344137; PubMed=9679194;  
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,  
Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,  
Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohiku Y.,  
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,  
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,  
Masuchi Y., Shizuya H., Kikuchi H.;  
RT "Complete sequence and gene organization of the genome of a hyper-  
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";  
RL DNA Res. 5:55-76(1998).  
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO  
FORM THE FILAMENTS OF FLAGELLA (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL FLAGELLIN FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
CC  
CC EMBL; AP000002; BAA29640.1; -.  
DR InterPro; IPR002774; Arch\_flagellin.  
DR Pfam; PF01917; Arch\_flagellin; 1.  
KW Flagella; Multigene family; Complete proteome.  
FT PROPEP 1 4 BY SIMILARITY.  
FT CHAIN 5 255 FLAGELLIN B5.  
SQ SEQUENCE 255 AA; 27311 MW; F33C04889F551832 CRC64;  
Query Match 72.1%; Score 31; DB 1; Length 255;  
Best Local Similarity 62.5%; Pred. No. 57;  
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 1 POGGKVT 8  
111111  
Db 67 PPGGKVT 74  
RESULT 15  
CAC4\_STRPN  
ID CAC4\_STRPN STANDARD; PRT; 299 AA.  
AC P58313;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE UDP--glucose-1-phosphate uridylyltransferase (EC 2.7.7.9) (UDP-glucose  
DE pyrophosphorylase) (UDPGP) (Alpha-D-glucosyl-1-phosphate  
DE uridylyltransferase) (uridine diphosphoglucose pyrophosphorylase).  
GN CAP4C OR SF2092.  
OS Streptococcus pneumoniae.  
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;  
OC Streptococcus.

```
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Pettelin H., Nelson K.E., Paulsen I.T., Elsen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Unayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Ufferback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -!- CATABOLIC ACTIVITY: GTP + alpha-D-glucose 1-phosphate =
CC diphosphate + UDP-glucose.
CC -!- PATHWAY: PRODUCTION OF CAPSULAR POLYSACCHARIDE.
CC -!- SIMILARITY: BELONGS TO THE PROKARYOTIC UDPGP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AE007498; AKK76151.1; -.
DR TIGR; SP2092; -.
DR InterPro; IPR001825; NTP_transferase.
KW Transferase; Kinase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 299 AA; 33278 MW; E9D644167B4356C3 CRC64;
-----
Query Match 72.1%; Score 31; DB 1; Length 299;
Best Local Similarity 83.3%; Pred. NO. 67;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
OY 1 POGOK 6
    |||..||
Db 179 POGOK 184
```

Search completed: October 28, 2002, 17:21:39  
Job time : 7.89474 secs

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:20 ; Search time 18.5263 Seconds  
(without alignments)  
74.702 Million cell updates/sec

Title: US-09-833-079-1  
Perfect score: 43  
Sequence: 1 POGGKVT 8

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues  
al number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

SPREMBL\_19: \*  
1: sp.archaea: \*  
2: sp.bacteria: \*  
3: sp.fungi: \*  
4: sp.human: \*  
5: sp.invertebrate: \*  
6: sp.mammal: \*  
7: sp.mhc: \*  
8: sp.organelle: \*  
9: sp.phage: \*  
10: sp.plant: \*  
11: sp.podent: \*  
12: sp.virus: \*  
13: sp.vertebrate: \*  
14: sp.unclassified: \*  
15: sp.virus: \*  
16: sp.bacteriap: \*  
17: sp.archaeap: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	43	100.0	30	09R530	Q9R530 escherichia
2	43	100.0	137	09R8B2	Q9R8B2 escherichia
3	43	100.0	163	09FD15	Q9FD15 escherichia
4	43	100.0	167	P71221	P71221 escherichia
5	43	100.0	167	09KHx1	Q9KHx1 escherichia
6	43	100.0	167	09KHw8	Q9KHw8 escherichia
7	43	100.0	171	09AM10	Q9AM10 escherichia
8	43	100.0	172	09L6D8	Q9L6D8 escherichia
9	43	100.0	174	09R3S9	Q9R3S9 escherichia
10	43	100.0	176	09K3C6	Q9K3C6 escherichia
11	43	100.0	176	P71222	P71222 escherichia
12	43	100.0	176	09KHx4	Q9KHx4 escherichia
13	43	100.0	176	09KHx3	Q9KHx3 escherichia
14	43	100.0	177	09KHx6	Q9KHx6 escherichia
15	43	100.0	177	09KHx2	Q9KHx2 escherichia
16	43	100.0	178	09R616	Q9R616 escherichia

17	43	100.0	183	2	099QA6	Q99QA6 escherichia
18	43	100.0	183	2	09A1L3	Q9A1L3 escherichia
19	43	100.0	187	2	P71217	P71217 escherichia
20	43	100.0	194	2	099Q97	Q99Q97 escherichia
21	43	100.0	194	2	09L6D9	Q9L6D9 escherichia
22	42	97.7	168	2	09R3R7	Q9R3R7 escherichia
23	40	93.0	159	2	09AM11	Q9AM11 escherichia
24	40	93.0	163	2	09FD16	Q9FD16 escherichia
25	40	93.0	171	2	09KHx0	Q9KHx0 escherichia
26	40	93.0	172	2	09L617	Q9L617 escherichia
27	40	93.0	172	2	09L6D7	Q9L6D7 escherichia
28	40	93.0	172	2	09L6D4	Q9L6D4 escherichia
29	40	93.0	172	2	09KHx5	Q9KHx5 escherichia
30	40	93.0	181	2	09KGS1	Q9KGS1 escherichia
31	40	93.0	182	2	09S417	Q9S417 escherichia
32	40	93.0	182	2	047195	Q47195 escherichia
33	39	90.7	168	2	P71218	P71218 escherichia
34	39	90.7	178	2	P71219	P71219 escherichia
35	39	90.7	178	2	09KHx7	Q9KHx7 escherichia
36	39	90.7	178	2	09KHw9	Q9KHw9 escherichia
37	36	83.7	162	2	09AM12	Q9AM12 escherichia
38	35	81.4	14	2	09R517	Q9R517 escherichia
39	35	81.4	187	2	Q47442	Q47442 escherichia
40	35	81.4	312	4	09BXU2	Q9BXU2 homo sapien
41	35	81.4	469	16	P74685	P74685 synechocyst
42	35	81.4	613	4	09UPB1	Q9UPB1 homo sapien
43	34	79.1	134	10	09M2N2	Q9M2N2 arabidopsis
44	34	79.1	745	16	Q32144	Q32144 bacillus su
45	34	79.1	1620	4	Q9ULT8	Q9ULT8 homo sapien

#### ALIGNMENTS

RESULT 1  
Q9R530 PRELIMINARY; PRT; 30 AA.  
AC Q9R530:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)  
DE 15.5 KDA FIMBRIAL SUBUNIT (FRAGMENT).  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.  
CC Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE.  
RX MEDLINE=94042915; PubMed=7901197;  
RA Giron J.A., Ho A.S., Schoolnik G.K.;  
RT "Characterization of fimbriae produced by enteropathogenic Escherichia  
RT coli."  
RL J. Bacteriol. 175:7391-7403(1993).  
SQ SEQUENCE 30 AA; 2964 MW; 5F63B427E63FFA90 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. NO. 0.15; Mismatches 0; Indels 0;  
Matches 8; Conservative 0; Mismatches 0; Gaps 0;

QY 1 POGGKVT 8  
| | | | | | | |  
Db 2 POGGKVT 9  
  
RESULT 2  
Q9R8B2 PRELIMINARY; PRT; 137 AA.  
AC Q9R8B2:  
DT 01-MAY-2000 (TREMBLrel. 13, Created)  
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE PAPA (FRAGMENT).  
GN PAPA.

OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CFT073;  
 RX MEDLINE=97342757; PubMed=9199454;  
 RA Kao J.S., Stucker D.M., Warren J.W., Mobley H.L.;  
 RT "Pathogenicity island sequences of *Yersinia enterocolitica* serotype 4/O:3 are associated with virulent uropathogenic strains.";  
 RL Infect. Immun. 65:2812-2820(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-CFT073;  
 RX MEDLINE=98380392; PubMed=9712795;  
 RA Guyer D.M., Kao J.S., Mobley H.L.T.;  
 RT "Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter associated bacteriuria and from fecal samples.";  
 RL Infect. Immun. 66:4411-4417(1998).  
 DR EMBL: AF081285; AAC61718.1; -;  
 DR InterPro: IPR000259; Fimbrin.  
 DR Pfam: PF00419; Fimbrin; 1.  
 FT NON\_TER 137 137  
 SQ SEQUENCE 137 AA; 13906 MW; 25F6B5BCEP5E9D7 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 137;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
 Db 26 POGGKVT 33

RESULT 3  
 O9FD15 PRELIMINARY; PRT; 163 AA.  
 ID O9FD15;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE PAPA VARIANT F48 (FRAGMENT).  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-3084;  
 RA Johnson J.R., Stell A.L.;  
 RT "The papa variant F48 in previously F-type Negative *Escherichia coli* strains.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF250040; AAG00915.1; -;  
 FT NON\_TER 1 1  
 SQ SEQUENCE 163 AA; 16727 MW; B5BBECA87A115908 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 163;  
 Best Local Similarity 100.0%; Pred. No. 0.89;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
 Db 6 POGGKVT 13

RESULT 4  
 O9KH21 PRELIMINARY; PRT; 167 AA.  
 ID P71221  
 AC P71221;

DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE F16 P-FIMBRIAE MAJOR SUBUNIT PRECURSOR.  
 GN SIA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC Escherichia.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-C83-83;  
 RX MEDLINE=20143777; PubMed=10678978;  
 RA Johnson J.R., Stell A.L., Schuetz F., O'Bryan T.T., Russo T.A.,  
 RA Carlinson U.B., Fasching C., Kavie J., Van Dijk L., Gastra W.;  
 RT "Analysis of the F antigen-specific papa alleles of extraintestinal *RT* pathogenic *Escherichia coli* using a novel multiplex PCR-based assay.";  
 RL Infect. Immun. 68:1587-1599(2000).  
 DR EMBL: Y08930; CAA70134.1; -;  
 DR InterPro: IPR000259; Fimbrin.  
 DR Pfam: PF00419; Fimbrin; 1.  
 DR Signal.  
 FT CHAIN 1 10 POTENTIAL.  
 FT CHAIN 11 167 POTENTIAL.  
 SQ SEQUENCE 167 AA; 16987 MW; 3CCIC0E061C00533 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 0.91;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
 Db 12 POGGKVT 19

RESULT 5  
 O9KH1 PRELIMINARY; PRT; 167 AA.  
 ID O9KH1;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE ADHESION PROTEIN PAPA.  
 OS Escherichia coli.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ECOR 60;  
 RX MEDLINE=21062863; PubMed=11106538;  
 RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
 RT "Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*.";  
 RL J. Infect. Dis. 185:78-88(2001).  
 DR EMBL: AF247354; AAF97911.1; -;  
 SQ SEQUENCE 167 AA; 17056 MW; 4BAE7ECC23D1E095 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 167;  
 Best Local Similarity 100.0%; Pred. No. 0.91;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 POGGKVT 8  
 Db 12 POGGKVT 19

RESULT 6  
 O9KH8 PRELIMINARY; PRT; 167 AA.  
 ID O9KH8;  
 AC O9KH8;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)



```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ADHESION PROTEIN PAPA.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR 66;
RA JEDLINE=21062863; PubMed=11106538;
RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;
RT "Phylogenetic distribution of extraintestinal virulence-associated
RT traits in Escherichia coli."
RL Infect. Dis. 183:78-86(2001).
DR EMBL; AF247357; AAF97914.1; -.
SQ SEQUENCE 167 AA; 17012 MW; 3BC3CEBC2684F4C5 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.91;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 POGGKVT 8
Db 12 POGGKVT 19
|||||

RESULT 7
ID 09AM10 PRELIMINARY; PRT; 171 AA.
AC 09AM10;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAPA (FRAGMENT).
GN PAPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
RN [1]
RP NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BOS020;
RA MEDLINE=21391793; PubMed=11500406;
RA Johnson J.R., O'Bryan T.T., Kuskowski M., Maslow J.N.;
RT "Ongoing Horizontal and Vertical Transmission of Virulence Genes and
RT PAPA Alleles among Escherichia coli Blood Isolates from Patients with
RT Diverse-Source Bacteremia."
RL Infect. Immun. 69:5363-5374(2001).
DR EMBL; AF332520; AAK11634.1; -.
FT NON-TER 1
SQ SEQUENCE 171 AA; 17540 MW; FE9653A1C7C39773 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 POGGKVT 8
Db 3 POGGKVT 10
|||||

RESULT 8
ID 09L6D8 PRELIMINARY; PRT; 172 AA.
AC 09L6D8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PAPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
RN [1]
RP NCBI_TaxID=562;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=5275(20); PubMed=10816481;
RX MEDLINE=20278115; PubMed=10816481;
RA Johnson J.R., O'Bryan T.T., Low D.A., Ling G., Delavari P.,
RA Fasching C., Russo T.A., Carlini U., Stell A.L.;
RT "Evidence of commonality between canine and human extraintestinal
RT pathogenic Escherichia coli strains that express paps allele III."
RL Infect. Immun. 68:3327-3336(2000).
DR EMBL; AF237478; AAF1959.1; -.
SQ SEQUENCE 172 AA; 17570 MW; 11B2438630A40E9D CRC64;

Query Match 100.0%; Score 43; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 0.94;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 POGGKVT 8
Db 15 POGGKVT 22
|||||

RESULT 9
ID 09R3S9 PRELIMINARY; PRT; 174 AA.
AC 09R3S9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR PILIN PROTEIN PRECURSOR (FRAGMENT).
GN PAPA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
RN [1]
RP NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECOR 53, ECOR 48, AND ECOR 52;
RX MEDLINE=98404145; PubMed=9732452;
RA Boyd E.F., Hartl D.L.;
RT "Diversifying selection governs sequence polymorphism in the major
RT adhesin proteins fliA, papa, and sfa of Escherichia coli."
RL J. Mol. Evol. 47:258-267(1998).
DR EMBL; AF051815; AAD02666.1; -.
DR EMBL; AF051811; AAD02668.1; -.
DR EMBL; AF051814; AAD02664.1; -.
DR InterPro; IPR000259; Fimbrin.
DR Pfam; PF00419; Fimbrin; 1.
FT NON-TER 1
SQ SEQUENCE 174 AA; 17628 MW; 4898EF59F7413839 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 0.95;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 POGGKVT 8
Db 16 POGGKVT 23
|||||

RESULT 10
ID 09K3C6 PRELIMINARY; PRT; 176 AA.
AC 09K3C6;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ADHESION PROTEIN PAPA (PAPA VARIANT F40).
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX Escherichia.
RN [1]
RP NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.

```

RC STRAIN-ECOR 41, Ecor 39, AND Ecor 40;  
RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
RT "Phylogenetic Distribution of Extraintestinal Virulence-Associated  
Traits in *Escherichia coli*."  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-V29;  
RA Johnson J.R., Stell A.L., Scheutz F., O'Bryan T.T., Russo T.A.,  
Carlino U.B., Fasching C., Van Dijk L., Gastra W.;  
RT "Analysis of the F Antigen-Specific PAPA Alleles of Extraintestinal  
Pathogenic *Escherichia coli* Using a Novel Multiplex PCR-Based Assay";  
RL Infect. Immun. 68:0-0(2000).  
DR EMBL; AF247352; AAF97909.1; -;  
DR EMBL; AF234627; AAF40211.1; -;  
DR EMBL; AF247350; AAF97907.1; -;  
DR EMBL; AF247351; AAF97908.1; -;  
DR InterPro: IPR000259; FimDrial.  
DR Pfam: PF00419; FimDrial; 1.  
SQ SEQUENCE 176 AA; 17780 MW; 0B4FDF3B0CCB37FE CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 POGGKVT 8  
Db 15 POGGKVT 22  
|||||  
RESULT 11  
ID P71222 PRELIMINARY; PRT; 176 AA.  
AC P71222;  
DT 01-FEB-1997 (Tremblrel. 02, Created)  
DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE F10 P-FIMBRIAE MAJOR SUBUNIT PRECURSOR.  
GN FTEA.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-C1960-79;  
MEDLINE=20143777; PubMed=10678978;  
RA Johnson J.R., Stell A.L., Scheutz F., O'Bryan T.T., Russo T.A.,  
Carlino U.B., Fasching C., Kave J., Van Dijk L., Gastra W.;  
RT "Analysis of the F antigen-specific PAPA alleles of extraintestinal  
pathogenic *Escherichia coli* using a novel multiplex PCR-based assay";  
RL Infect. Immun. 68:1587-1599(2000).  
DR EMBL; Y08927; CAA70131.1; -;  
DR InterPro: IPR000259; FimDrial.  
DR Pfam: PF00419; FimDrial; 1.  
KW Signal.  
FT SIGNAL.  
FT CHAIN 11 176 POTENTIAL.  
SQ SEQUENCE 176 AA; 17686 MW; A2B033F7F578AD1F CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 POGGKVT 8  
Db 15 POGGKVT 22  
|||||  
RESULT 12  
ID Q9KH4 PRELIMINARY; PRT; 176 AA.  
AC Q9KH4;

DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE ADHESION PROTEIN PAPA.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ECOR 35;  
MEDLINE=21062863; PubMed=11106538;  
RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
RT "Phylogenetic distribution of extraintestinal virulence-associated  
traits in *Escherichia coli*."  
RL Infect. Dis. 183:78-88(2001).  
DR EMBL; AF247348; AAF97905.1; -;  
SQ SEQUENCE 176 AA; 17860 MW; 51DD43CA682D40E0 CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 POGGKVT 8  
Db 15 POGGKVT 22  
|||||  
RESULT 13  
ID Q9KH3 PRELIMINARY; PRT; 176 AA.  
AC Q9KH3;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE ADHESION PROTEIN PAPA.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-ECOR 36;  
MEDLINE=21062863; PubMed=11106538;  
RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
RT "Phylogenetic distribution of extraintestinal virulence-associated  
traits in *Escherichia coli*."  
RL Infect. Dis. 183:78-88(2001).  
DR EMBL; AF247349; AAF97906.1; -;  
SQ SEQUENCE 176 AA; 17860 MW; 148917C9DFCE657 CRC64;  
Query Match 100.0%; Score 43; DB 2; Length 176;  
Best Local Similarity 100.0%; Pred. No. 0.97;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 POGGKVT 8  
Db 15 POGGKVT 22  
|||||  
RESULT 14  
ID Q9KH6 PRELIMINARY; PRT; 177 AA.  
AC Q9KH6;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE ADHESION PROTEIN PAPA.  
OS *Escherichia coli*.  
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
OC *Escherichia*.  
OX NCBI\_TaxID=562;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN-ECOR 24;  
 RX MEDLINE-21062863; PubMed-11106538;  
 RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
 RT "Phylogenetic distribution of extraintestinal virulence-associated  
 traits in *Escherichia coli*."  
 RL J. Infect. Dis. 183:78-88(2001).  
 DR EMBL: AF247346; AAF97903.1;  
 SQ SEQUENCE 177 AA; 18171 MW; 822C14624B5DE8C1 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 0.97;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 15 POGGKVT 22

## RESULT 15

Q9KH2 PRELIMINARY; PRT; 177 AA.  
 Q9KH2;  
 DT 01-OCT-2000 (TREMblrel. 15, Created)  
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)  
 DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)  
 DE ADHESION PROTEIN PAPA.  
 OS *Escherichia coli*.  
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
 OC *Escherichia*.  
 OX NCBI\_TaxID=562;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-ECOR 56;  
 RX MEDLINE-21062863; PubMed-11106538;  
 RA Johnson J.R., Delavari P., Kuskowski M., Stell A.L.;  
 RT "Phylogenetic distribution of extraintestinal virulence-associated  
 traits in *Escherichia coli*."  
 RL J. Infect. Dis. 183:78-88(2001).  
 DR EMBL: AF247353; AAF97910.1;  
 SQ SEQUENCE 177 AA; 18143 MW; E95C03D49033ADD5 CRC64;

Query Match 100.0%; Score 43; DB 2; Length 177;  
 Best Local Similarity 100.0%; Pred. No. 0.97;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 POGGKVT 8  
 |||||  
 Db 15 POGGKVT 22

Search completed: October 28, 2002, 17:24:03  
 Job time : 20.5263 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2002, 17:21:15 ; Search time 30.6842 Seconds  
(without alignments)  
39.819 Million cell updates/sec

Title: US-09-833-079-2  
Perfect score: 58  
Sequence: 1 AKFGCMGAKKG 11

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

al number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : A-Geneseq\_032802:\*

1:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22:	/SIDSL/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	39	67.2	449	ABG09511
2	39	67.2	455	ABG08528
3	39	67.2	455	ABG09503
4	39	67.2	536	ABG12830
5	38	65.5	126	AA19197
6	38	65.5	132	AA19199
7	38	65.5	136	AA19196
8	38	65.5	21	AA190088
9	37	63.8	574	ABB67213
10	37	63.8	577	ABB64981
11	37	63.8	578	ABB65733

12	37	63.8	578	22	ABB67212	Drosophila melanog
13	37	63.8	989	22	AA90769	Human shear stress
14	36	62.1	71	22	AA009104	Human polypeptide
15	36	62.1	131	21	AA19198	Amino acid sequenc
16	36	62.1	137	22	AA000939	Human polypeptide
17	36	62.1	142	22	AA000501	Human polypeptide
18	36	62.1	163	6	AA050045	Sequence of the ga
19	36	62.1	163	7	AA060247	E.coli H0849 gal-G
20	36	62.1	163	22	AA047081	pillin protein, Pap
21	36	62.1	163	22	AA072835	Bacterial P pilus
22	36	62.1	198	22	AA088679	Human Immune/haema
23	36	62.1	388	19	AA070505	Pyrococcus horikos
24	36	62.1	576	22	AA089928	C glutamicum prote
25	36	62.1	637	22	AA079357	Corynebacterium g1
26	36	62.1	886	22	AA071328	Drosophila melanog
27	35	60.3	113	22	AB067899	Drosophila melanog
28	35	60.3	116	21	AA016428	EucaIyplus grandis
29	35	60.3	221	22	AA070637	POBRV envelope pro
30	35	60.3	648	22	AB059450	Drosophila melanog
31	35	60.3	652	22	AB027471	Novel human diagno
32	35	60.3	657	18	AA032095	Porcine retrovirus
33	35	60.3	660	20	AA085453	Pig endogenous ret
34	35	60.3	660	22	AA070632	Porcine endogenous
35	35	60.3	660	22	AA035113	PERV-A env protein
36	35	60.3	809	22	AA041064	Human polypeptide
37	35	60.3	872	22	AA086163	Human MTR1 protein
38	35	60.3	1069	22	AA073284	Defective retrovir
39	35	60.3	1158	22	AA086164	Mouse MTR1 protein
40	35	60.3	1158	22	AA086166	Mouse MTR1 protein
41	35	60.3	1156	22	AA086162	Human MTR1 protein
42	35	60.3	1179	22	AA086165	Human MTR1 protein
43	35	60.3	1189	22	AB059904	Drosophila melanog
44	35	60.3	10	22	AA084320	Arabidopsis thalia
45	34	58.6	79.	21	AA08378	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
ABG09511  
ID ABC09511 standard; Protein; 449 AA.  
XX  
AC ABG09511;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #9502.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RF, Liu C, Tang YT;  
XX  
XX WPI: 2001-639362/73.  
XX  
XX N-PSDB; AAS73698.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess

PT biodiversity -  
 XX Claim 20; SEQ ID No 39870; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 449 AA;

Query Match 67.2%; Score 39; DB 22; Length 449;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 AKFGMGAKKG 11  
 | | | | | : | |  
 Db 327 ADFGSGSTRKG 337

RESULT 2  
 ABG08528  
 ID ABG08528 standard; Protein; 455 AA.  
 XX  
 AC ABG08528;  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #8519.  
 XX  
 OS Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 OS food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS72715.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 39870; 103pp; English.

PS Claim 20; SEQ ID No 38867; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 455 AA;

Query Match 67.2%; Score 39; DB 22; Length 455;  
 Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 AKFGMGAKKG 11  
 | | | | | : | |  
 Db 191 ADFGSGSTRKG 201

RESULT 3  
 ABG09503  
 ID ABG09503 standard; Protein; 455 AA.  
 XX  
 AC ABG09503;  
 DT 13-FEB-2002 (first entry)  
 XX  
 DE Novel human diagnostic protein #9494.  
 XX  
 OS Human: chromosome mapping; gene mapping; gene therapy; forensic;  
 OS food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.  
 XX WO200175067-A2.  
 XX 11-OCT-2001.  
 XX 30-MAR-2001; 2001WO-US08631.  
 XX 31-MAR-2000; 2000US-0540217.  
 XX 23-AUG-2000; 2000US-0649167.  
 XX (HYSE-) HYSEQ INC.  
 XX Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 DR N-PSDB; AAS73690.  
 XX New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity -  
 XX Claim 20; SEQ ID No 39862; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 455 AA;  
Query Match 67.2%; Score 39; DB 22; Length 455;  
Best Local Similarity 63.6%; Pred. No. 1.2e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AKFGMGAKKG 11  
| | | | | : | |  
Db 191 ADFGSGSTRKG 201

RESULT 4  
ABG12830  
ID ABG12830 standard; Protein; 536 AA.  
XX  
AC ABG12830;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #12821.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
WO200175067-A2.  
XX  
11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR N-PSDB: AAS77017.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 20; SEQ ID NO 43189; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pcl\_sequences.

SQ Sequence 536 AA;  
Query Match 67.2%; Score 39; DB 22; Length 536;  
Best Local Similarity 63.6%; Pred. No. 1.4e+02;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
QY 1 AKFGMGAKKG 11  
| | | | | : | |  
Db 272 ADFGSGSTRKG 282

RESULT 5  
AAB19197  
ID AAB19197 standard; Protein; 126 AA.  
XX  
AC AAB19197;  
XX  
DT 19-FEB-2001 (first entry)  
XX  
DE Amino acid sequence of a bay scallop abductin polypeptide.  
XX  
KW Bay scallop; abductin; chemomechanical transduction; drug delivery;  
KW inverse temperature transition; water soluble drug; biomaterial;  
XX  
KW fabric; organ prosthesis.  
XX  
OS Argopecten sp.  
XX  
OS  
XX US6127166-A.  
XX  
PN  
XX 03-OCT-2000.  
XX  
PD 03-NOV-1997; 97US-0963168.  
XX  
PE 03-NOV-1997; 97US-0963168.  
XX  
PR 03-NOV-1997; 97US-0963168.  
XX  
PA (BAYL/) BAYLEY H.  
PA (CAOQ/) CAO Q.  
PA (WANG/) WANG Y.  
XX  
XX Bayley H, Cao Q, Wang Y;  
XX  
PI  
XX  
DR WPI: 2000-611057/58.  
DR N-PSDB: AANC61378.  
XX  
XX Abductin nucleic acid molecules, useful for expressing abductin  
PT polypeptides which are used in the manufacture of drug delivery  
PT vehicles for administering water soluble drugs -  
XX  
PS Example 1; Fig 2; 30pp; English.  
XX  
XX The present sequence represents a bay scallop abductin polypeptide.  
CC The polypeptide is capable of chemomechanical transduction or inverse  
CC temperature transition. Abductin polypeptides contain glycine-rich

CC sequences. Abductin polypeptides are useful in the manufacture of  
 CC drug delivery vehicles for administering water soluble drugs. The  
 CC abductin polypeptides and their derivatives are also useful in the  
 CC manufacture of broad range of biomaterials ranging from light-weight  
 CC durable fabric for clothing to matrices useful for human tissue and  
 CC organ prostheses.

XX Sequence 126 AA;

Query Match 65.5%; Score 38; DB 21; Length 126;

Best Local Similarity 77.8%; Pred. No. 56;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 FGGMGAKKG 11

Db 82 FGGMGCGKG 90

RESULT 6

9199

AA19199 standard; Protein; 132 AA.

AA19199;

19-FEB-2001 (first entry)

Amino acid sequence of a bay scallop abductin polypeptide.

Bay scallop; abductin; chemomechanical transduction; drug delivery;

Inverse temperature transition; water soluble drug; biomaterial;

fabric; organ prostheses.

Argopecten sp.

US6127166-A.

03-OCT-2000.

03-NOV-1997; 97US-0963168.

03-NOV-1997; 97US-0963168.

(BAYL/) BAYLEY H.

(CAOQ/) CAO Q.

(WANG/) WANG Y.

Bayley H, Cao Q, Wang Y;

WPI; 2000-611057/58.

N-PSDB; AAC61380.

Abductin nucleic acid molecules, useful for expressing abductin

polypeptides which are used in the manufacture of drug delivery

vehicles for administering water soluble drugs -

Example 1; Fig 2; 30pp; English.

The present sequence represents a bay scallop abductin polypeptide.

The polypeptide is capable of chemomechanical transduction or inverse

temperature transition. Abductin polypeptides contain glycine-rich

sequences. Abductin polypeptides are useful in the manufacture of

drug delivery vehicles for administering water soluble drugs. The

abductin polypeptides and their derivatives are also useful in the

manufacture of broad range of biomaterials ranging from light-weight

QY 3 FGGMGAKKG 11

Db 82 FGGMGCGKG 90

RESULT 7

AA19196

AA19196 standard; Protein; 136 AA.

AA19196;

19-FEB-2001 (first entry)

Amino acid sequence of a bay scallop abductin polypeptide.

Bay scallop; abductin; chemomechanical transduction; drug delivery;

Inverse temperature transition; water soluble drug; biomaterial;

fabric; organ prostheses.

Argopecten sp.

US6127166-A.

03-OCT-2000.

03-NOV-1997; 97US-0963168.

03-NOV-1997; 97US-0963168.

(BAYL/) BAYLEY H.

(CAOQ/) CAO Q.

(WANG/) WANG Y.

Bayley H, Cao Q, Wang Y;

WPI; 2000-611057/58.

N-PSDB; AAC61376, AAC61377.

Abductin nucleic acid molecules, useful for expressing abductin

polypeptides which are used in the manufacture of drug delivery

vehicles for administering water soluble drugs -

Example 1; Fig 2; 30pp; English.

The present sequence represents a bay scallop abductin polypeptide.

The polypeptide is capable of chemomechanical transduction or inverse

temperature transition. Abductin polypeptides contain glycine-rich

sequences. Abductin polypeptides are useful in the manufacture of

drug delivery vehicles for administering water soluble drugs. The

abductin polypeptides and their derivatives are also useful in the

manufacture of broad range of biomaterials ranging from light-weight

durable fabric for clothing to matrices useful for human tissue and

organ prostheses.

Example 1; Fig 2; 30pp; English.

The present sequence represents a bay scallop abductin polypeptide.

The polypeptide is capable of chemomechanical transduction or inverse

temperature transition. Abductin polypeptides contain glycine-rich

sequences. Abductin polypeptides are useful in the manufacture of

QY 3 FGGMGAKKG 11

Db 82 FGGMGCGKG 90

RESULT 8

AA19196

AA19196 standard; Protein; 136 AA.

AA19196;

19-FEB-2001 (first entry)

Amino acid sequence of a bay scallop abductin polypeptide.

Bay scallop; abductin; chemomechanical transduction; drug delivery;

Inverse temperature transition; water soluble drug; biomaterial;

fabric; organ prostheses.

Argopecten sp.

US6127166-A.

03-OCT-2000.

03-NOV-1997; 97US-0963168.

03-NOV-1997; 97US-0963168.

(BAYL/) BAYLEY H.

(CAOQ/) CAO Q.

(WANG/) WANG Y.

Bayley H, Cao Q, Wang Y;

WPI; 2000-611057/58.

N-PSDB; AAC61376, AAC61377.

Abductin nucleic acid molecules, useful for expressing abductin

polypeptides which are used in the manufacture of drug delivery

vehicles for administering water soluble drugs -

Example 1; Fig 2; 30pp; English.

The present sequence represents a bay scallop abductin polypeptide.

The polypeptide is capable of chemomechanical transduction or inverse

temperature transition. Abductin polypeptides contain glycine-rich

sequences. Abductin polypeptides are useful in the manufacture of

drug delivery vehicles for administering water soluble drugs. The

abductin polypeptides and their derivatives are also useful in the

manufacture of broad range of biomaterials ranging from light-weight

durable fabric for clothing to matrices useful for human tissue and

organ prostheses.

Example 1; Fig 2; 30pp; English.

The present sequence represents a bay scallop abductin polypeptide.



DE Aldehyde dehydrogenase fccd amino acid sequence SEQ ID NO:9.  
 XX  
 KW Bacterium 2412.1; maize; detoxification; degradation; carcinogen;  
 XX fumonisin catabolic gene cluster; antifungal; mycotoxin.  
 OS Bacterium 2412.1.  
 XX  
 PN WO200004158-A2.  
 XX  
 PD 27-JAN-2000.  
 XX  
 PF 14-JUL-1999; 99WO-US15837.  
 XX  
 PR 15-JUL-1998; 98US-0092953.  
 XX  
 PA (PION-) PIONEER HI-BRED INT INC.  
 XX  
 PI Duvick JP, Maddox J, Gilliam J;  
 XX  
 DR WPI: 2000-171270/15.  
 N-PSDB; AA291257.  
 PS  
 PT New isolated nucleic acid molecules which encode fumonisin degradative  
 PT enzymes and transporters, used for detoxifying fumonisin or related  
 PT mycotoxin or reducing the pathogenicity of fungi producing fumonisin -  
 XX  
 PS Claim 1; Page 106-108; 181pp; English.  
 XX  
 CC The present invention describes isolated nucleic acid molecules which  
 CC encode fumonisin degradative enzymes and transporters from a  
 CC Bacterium 2412.1 fumonisin catabolic gene cluster. The Bacterium 2412.1  
 CC was isolated from maize. The nucleic acid molecules can be used for  
 CC producing an enzyme for detoxifying fumonisin or a structurally related  
 CC mycotoxin, e.g. in harvested grain, or processed grain which is to be  
 CC used as animal feed or silage. Ruminant microorganisms transformed with  
 CC the nucleic acid molecules can be used as probiotic compositions or as  
 CC feed inoculant compositions. Plants transformed with the nucleic acid  
 CC molecules can degrade or transport fumonisin and can reduce the  
 CC pathogenicity of a fungus producing fumonisin. The plants may be e.g.  
 CC maize, sorghum, wheat, tomato, soybean, alfalfa, sunflower, brassica,  
 CC cotton or rice. The products can also be used as detection reagents for  
 CC fumonins and related compounds. AA291253 represents the Bacterium  
 CC 2412.1 fumonisin catabolic gene cluster, and AA291254 to AA291273  
 CC represent isolated nucleic acids from Bacterium 2412.1 which encode  
 CC AA291285 to AA291286 to AA291291 represent sequence used in the  
 CC exemplification of the present invention. AA291274 to AA291285 are given  
 CC in the sequence listing from the present specification but are not  
 CC specifically mentioned further.

Sequence 468 AA:  
 Query Match 65.5%; Score 38; DB 21; Length 468;  
 Best Local Similarity 63.6%; Pred. No. 1.9e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 AKFGGMAKKG 11  
 ||| |||::|  
 DB 441 AKFGSGSELG 451

RESULT 9  
 ABB67213  
 ID ABB67213 standard; Protein; 574 AA.  
 XX  
 AC ABB67213;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 28431.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX

OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 XX  
 DR WPI: 2001-656660/75.  
 XX  
 DR N-PSDB; ABL11316.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 28431; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins  
 CC (AB57737-AB872072).  
 CC  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.  
 CC  
 SQ Sequence 574 AA:  
 Query Match 63.8%; Score 37; DB 22; Length 574;  
 Best Local Similarity 70.0%; Pred. No. 3.3e+02;  
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 KFGGMAKKG 11  
 :||| | |||  
 DB 541 KFGGSGFKKG 550

RESULT 10  
 ABB64981  
 ID ABB64981 standard; Protein; 577 AA.  
 XX  
 AC ABB64981;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 21735.  
 XX  
 KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PR 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 XX  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX

PI Venter JC, Adams M, Li PWD, Myers EW;  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL09084.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 21735; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 577 AA;  
XX  
Query Match 63.8%; Score 37; DB 22; Length 577;  
Best Local Similarity 70.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 2 KFGGMGAKKG 11  
: || | | | |  
Db 544 RFGGGGFRKG 553  
RESULT 11  
ABB57733  
ID ABB65733 standard; Protein: 578 AA.  
XX  
AC ABB65733;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 23991.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
PA Drosophila melanogaster.  
XX  
PI WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL09836.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 23991; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 578 AA;  
XX  
Query Match 63.8%; Score 37; DB 22; Length 578;  
Best Local Similarity 70.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
OY 2 KFGGMGAKKG 11  
: || | | | |  
Db 545 RFGGGGFRKG 554  
RESULT 12  
ABB67212  
ID ABB67212 standard; Protein: 578 AA.  
XX  
AC ABB67212;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 28428.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PE 23-MAR-2001; 2001WO-US09231.  
XX  
PR 23-MAR-2000; 2000US-191637P.  
PR 11-JUL-2000; 2000US-0614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
XX WPI: 2001-656860/75.  
DR N-PSDB; ABL11315.  
XX  
PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signalling and cell-cell  
PT interactions -  
XX  
PS Disclosure; SEQ ID NO 28428; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins  
CC (ABB57737-ABB72072).  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 578 AA;  
XX  
Query Match 63.8%; Score 37; DB 22; Length 578;

Best Local Similarity 70.0%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 KFGMGAKKG 11  
          :| | | | |  
Db 545 RFGGCGFKKG 554

## RESULT 13

AAB90769  
ID AAB90769 standard; Protein: 989 AA.

XX AAB90769;

DT 15-JUN-2001 (first entry)

XX Human shear stress-response protein SEQ ID NO: 38.

DE Human shear stress-response protein; vascular disease;

KM arteriosclerosis.

XX Homo sapiens.

PN W0200125427-A1.

PD 12-APR-2001.

PF 02-OCT-2000; 2000WO-JP06840.

PR 01-OCT-1999; 99JP-0280976.

PA (KYOW ) KYOMA HAKKO KOSYO KK.

PA (NOJI/) NOJIMA H.

PI Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;

PI Kuga T, Sekine S, Nakamura Y, Sugano S;

DR WPI: 2001-266308/27.

DR N-PSDB; AAH02892.

XX DNA sequences, proteins encoded by them and antibodies against them

PT useful in diagnosis and treatment of vascular disease caused by

PT arteriosclerosis -

XX The present invention provides the protein and coding sequences of a

CC number of human shear stress response proteins. These are useful in the

CC diagnosis, treatment and screening of vascular diseases caused by

CC arteriosclerosis, including heart failure, post-PTCA restenosis and

CC hypertension.

XX Sequence 989 AA;

## Query Match

Best Local Similarity 70.0%; Pred. No. 5.4e+02;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 2 KFGMGAKKG 11  
          :| | | | |  
Db 163 KVGNGSKKG 172

## RESULT 14

AAO09104  
ID AAO09104 standard; Protein: 71 AA.

XX AAO09104;

DT 06-NOV-2001 (first entry)

XX Human polypeptide SEQ ID NO 22996.

KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KM tissue growth factor; immunomodulatory; cancer; leukaemia;

OS nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

PN W0200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB; AAI89035.

XX Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX Claim 20; SEQ ID NO 22996; 1399pp + Sequence listing; English.

PS The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 71 AA;

## Query Match

Best Local Similarity 77.8%; Pred. No. 70;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 RFGMGAKKG 11  
          || | | | |  
Db 63 RFGGGLKKG 71

## RESULT 15

AAB19198  
ID AAB19198 standard; Protein: 131 AA.

XX AAB19198;

DT 19-FEB-2001 (first entry)

XX Amino acid sequence of a bay scallop abductin polypeptide.

KM Bay scallop; abductin; chemomechanical transduction; drug delivery;

KM inverse temperature transition; water soluble drug; biomaterial;

KM fabric; organ prostheses.

XX Argopecten sp.

OS US6127166-A.

XX

PD 03-OCT-2000.  
 XX  
 PF 03-NOV-1997; 97US-0963168.  
 XX  
 PR 03-NOV-1997; 97US-0963168.  
 XX

PA (BAYL/) BAYLEY H.  
 PA (CAOQ/) CAO Q.  
 PA (WANG/) WANG Y.  
 XX

PI Bayley H, Cao Q, Wang Y;  
 XX

DR WPI: 2000-611057/58.  
 DR N-PSDB; AAC61379.  
 XX

PT Abductin nucleic acid molecules, useful for expressing abductin  
 PT polypeptides which are used in the manufacture of drug delivery  
 PT vehicles for administering water soluble drugs -

Example 1; Fig 2; 30pp; English.

CC The present sequence represents a bay scallop abductin polypeptide.  
 CC The polypeptide is capable of chemomechanical transduction or inverse  
 CC temperature transition. Abductin polypeptides contain glycine-rich  
 CC sequences. Abductin polypeptides are useful in the manufacture of  
 CC drug delivery vehicles for administering water soluble drugs. The  
 CC abductin polypeptides and their derivatives are also useful in the  
 CC manufacture of broad range of biomaterials ranging from light-weight  
 CC durable fabric for clothing to matrices useful for human tissue and  
 CC organ prostheses.  
 CC

SQ Sequence 131 AA;

Query Match 62.1%; Score 36; DB 21; Length 131;  
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 3 FCGMGAKKG 11  
 |||||  
 Db 78 FCGMAAKKG 86

Search completed: October 28, 2002, 17:22:40  
 Job time : 32.6842 secs